

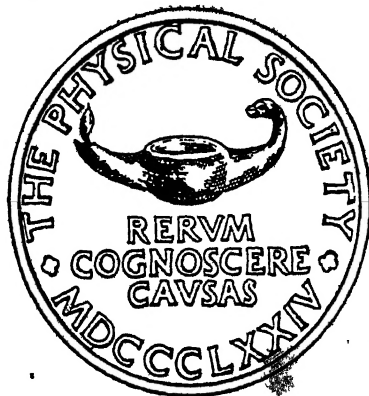
THE
PROCEEDINGS
OF
THE PHYSICAL SOCIETY

Section B

FROM JANUARY 1950 TO DECEMBER 1950

VOLUME 63

L53/95 -



48136
IARI

48136
c.17

Published by
THE PHYSICAL SOCIETY
1 Lowther Gardens, Prince Consort Road,
London S.W. 7

OFFICERS AND COUNCIL, 1950-51

PRESIDENT

L. F. BATES, D.Sc., Ph.D., F.R.S.

VICE-PRESIDENTS

who have filled the Office of President

C. H. LEES, D.Sc., F.R.S. (1918-20)
Sir FRANK SMITH, G.C.B., G.B.E., D.Sc., LL.D., F.R.S. (1924-26)
Sir OWEN RICHARDSON, M.A., D.Sc., F.R.S. (1926-28)
W. H. ECCLES, D.Sc., M.I.E.E., F.R.S. (1928-30)
A. O. RANKINE, O.B.E., D.Sc., F.R.S. (1932-34)
T. SMITH, M.A., F.R.S. (1936-38)
ALLAN FERGUSON, M.A., D.Sc. (1938-41)
Sir CHARLES DARWIN, K.B.E., M.C., M.A., Sc.D., F.R.S. (1941-43)
E. N. DA C. ANDRADE, Ph.D., D.Sc., F.R.S. (1943-45)
Sir DAVID BRUNT, M.A., Sc.D., F.R.S. (1945-47)
G. I. FINCH, M.B.E., D.Sc., F.R.S. (1947-49)
S. CHAPMAN, M.A., D.Sc., F.R.S. (1949-50)

VICE-PRESIDENTS

W. JEVONS, D.Sc., Ph.D. R. E. PEIERLS, C.B.E., M.A., D.Phil., D.S.
C. H. COLLIE, M.A., B.Sc. F.R.S.
D. ROAF, M.A., D.Phil.

HONORARY SECRETARIES

C. G. WYNNE, B.A. (*Business*) H. H. HOPKINS, Ph.D. (*Papers*)

HONORARY FOREIGN SECRETARY

E. N. DA C. ANDRADE, Ph.D., D.Sc., F.R.S.

HONORARY TREASURER

A. J. PHILPOT, C.B.E., M.A., B.Sc.

HONORARY LIBRARIAN

R. W. B. PRARSE, D.Sc., Ph.D.

ORDINARY MEMBERS OF COUNCIL

A. C. G. MENZIES, M.A., D.Sc.	A. B. WOOD, O.B.E., D.Sc.
J. H. AWBERY, M.A., B.Sc.	H. S. W. MASSEY, B.A., Ph.D., F.R.S.
R. C. EVANS, M.A., Ph.D.	L. J. DAVIES, M.A., B.Sc.
L. C. MARTIN, D.Sc.	W. V. MAYNEORD, D.Sc.
C. E. WYNN-WILLIAMS, D.Sc., Ph.D.	R. W. B. STEPHENS, B.Sc., Ph.D.
A. G. QUARRELL, D.Sc., Ph.D.	S. MARSH, D.Sc., Ph.D.

COLOUR GROUP

Chairman

W. S. STILES, O.B.E., D.Sc., Ph.D.

Honorary Secretary

R. G. HORNER, B.A.

LOW-TEMPERATURE GROUP

Chairman

F. E. SIMON, C.B.E., M.A., D.Phil.,
F.R.S.

Honorary Secretary

G. G. HASELDEN, Ph.D.

OPTICAL GROUP

Chairman

A. C. G. MENZIES, M.A., D.Sc.

Honorary Secretary

K. J. HABELL, M.Sc.

ACOUSTICS GROUP

Chairman

H. L. KIRKE, C.B.E., M.I.E.E.

Joint Honorary Secretaries

W. A. ALLEN, B.Arch., A.R.I.B.A.
A. T. PICKLES, O.B.E., M.A.

SECRETARY-EDITOR

Miss A. C. STICKLAND, M.Sc., Ph.D.

1 Lowther Gardens, Prince Consort Road, London S.W.7
(Telephone : KENsington 0048)

CONTENTS

Part 1. January 1950

	PAGE
Editorial	1
Mr. J. K. MACKENZIE. The Elastic Constants of a Solid containing Spherical Holes	2
Prof. FRIEDRICH BLAHA. On Movements of Small Ferromagnetic Particles in Inhomogeneous Magnetic Fields	12
Mr. B. COLLINGE. Dead Times of Self-Quenching Counters	15
Mr. F. G. HEYMANN. Breakdown in Cold-Cathode Tubes at Low Pressure	25
Mr. W. E. WILLSHAW and Mr. R. G. ROBERTSHAW. The Behaviour of Multiple Circuit Magnetrons in the Neighbourhood of the Critical Anode Voltage	41
Mr. D. G. KIELY. Measurements of the Reflection Coefficient of Water at a Wavelength of 8.7 mm.	46
Mr. T. L. ECKERSLEY. Coupling of the Ordinary and Extraordinary Rays in the Ionosphere	49
Letters to the Editor :	
Mr. A. N. HUNTER. The Debye Effect in Electrolytes	58
Mr. C. E. CHALLICE. Electron Optics of the Three-Stage Electron Microscope	59
Reviews of Books	62
Contents for Section A	63
Abstracts for Section A	63

Part 2. February 1950

Dr. J. B. BIRKS. The Properties of Ferromagnetic Compounds at Centimetre Wavelengths	65
Dr. L. JACOB. The Field in an Electron-Optical Immersion Objective	75
Miss M. GILBERT. Colour Perception in Parafoveal Vision	83
Dr. D. BROMLEY and Dr. R. H. HERZ. Quantum Efficiency in Photographic X-ray Exposures	90
Mr. B. H. BRIGGS, Mr. G. J. PHILLIPS and Mr. D. H. SHINN. The Analysis of Observations on Spaced Receivers of the Fading of Radio Signals	106
M. R. BUREAU. Les renforcements brusques des ondes très longues	122
Dr. R. RIVAUULT. Diffusion des échos au voisinage des fréquences critiques de F 2	126
Mr. D. R. BATES and Mr. M. J. SEATON. Theoretical Considerations Regarding the Formation of the Ionized Layers	129
Physical Society Conference at Cambridge—Abstracts	141
Reviews of Books	151
Contents for Section A	154
Abstracts for Section A	155

Part 3. March 1950

Dr. G. G. HASELDEN. The Deposition of Ice Crystals on Cooling Surfaces in Low Temperature Plant	157
Mr. T. S. MOSS. A Relationship between the Refractive Index and the Infra-Red Threshold of Sensitivity for Photoconductors	167
Dr. R. COOPER. Capacitance Measurements on Selenium Rectifiers: Evidence of Anomalous Dispersion	176
Mr. J. W. WARREN, Mr. W. HOPWOOD and Dr. J. D. CRAGGS. On Dissociation Processes in Certain Gases of High Dielectric Strength	180

	PAGE
Dr. R. C. PARKER and Mr. D. HATCH. The Static Coefficient of Friction and the Area of Contact	185
Prof. E. N. DA C. ANDRADE and Dr. R. F. Y. RANDALL. The Thermal Etching of Single Crystals of Cadmium	198
Mr. H. D. KEITH. The Lattice-Parameters of Clear Crystalline Quartz	208
Letters to the Editor :	
Dr. J. COLLARD, Mr. G. R. NICOLL and Dr. A. W. LINES. Discrepancies in the Measurement of Microwave Power at Wavelengths below 3 cm.	215
Reviews of Books	216
Corrigenda	221
Contents for Section A	222
Abstracts for Section A	222

Part 4. April 1950

Prof. A. O. RANKINE. Experimental Studies in Thermal Convection (33rd Guthrie Lecture)	225
Dr. G. M. B. DOBSON. Physics and the Atmosphere (5th Charles Chree Address)	252
Mr. B. G. CHILDS and Mr. S. WEINTROUB. The Measurement of the Thermal Expansion of Single Crystals of Tin by an Interferometric Method	267
Dr. O. BUNEMAN. A Toroidal Magnetron	278
Mr. C. N. DAVIES. Viscous Flow Transverse to a Circular Cylinder	288
Mr. F. H. KRENZ. On the Fluctuating Concentration of X-Ray Products in Water Dispersions	297
Dr. R. FÜRTH and Dr. D. K. C. MACDONALD. On the Retarding Field Current in Diodes	300
Contents for Section A	303
Abstracts for Section A	303

Part 5. May 1950

Mr. G. BRADFIELD. Summarized Proceedings of Symposium on Applications of Ultrasonics	305
Mr. A. C. LYNCH. Measurement of the Equivalent Electrical Circuit of a Piezoelectric Crystal	323
Dr. S. HARPER and Prof. A. H. COTTRELL. Surface Effects and the Plasticity of Zinc Crystals	331
Mr. H. L. WAIN and Prof. A. H. COTTRELL. Yield Points in Zinc Crystals	339
Mr. L. M. T. HOPKIN. A Simple Constant Stress Apparatus for Creep Testing.	346
Dr. G. L. J. BAILEY and Mr. H. C. WATKINS. Surface Tensions in the System Solid Copper-Molten Lead	350
Dr. K. M. GREENLAND and Mr. C. BILLINGTON. The Construction of Interference Filters for the Transmission of Light of Specified Wavelengths	355
Mr. D. R. BARBER. Note on the Brightness Profile and Photometric Contrast of a Test-object having small Angular Dimensions and Silhouetted against the Twilight Sky	364
Letters to the Editor :	
Dr. R. F. SAXE and Dr. J. B. HIGHAM. The Measurement of Spark Channel Diameters	370
Mr. H. A. GEBBIE, Mr. P. C. BANBURY and Dr. C. A. HOGARTH. Crystal Diode and Triode Action in Lead Sulphide	371
Reviews of Books	371
Contents for Section A	371
Abstracts for Section A	371

Contents

v

Part 6. June 1950

	PAGE
Dr. K. D. FROOME. The Behaviour of the Cathode Spot on an Undisturbed Liquid Surface of Low Work Function	377
Dr. MARY B. HESSE. The Calculation of Magnetic Lens Fields by Relaxation Methods	386
Dr. D. J. MALAN and Dr. B. F. J. SCHONLAND. An Electrostatic Fluxmeter of Short Response-time for use in Studies of Transient Field-changes	402
Dr. G. G. MACFARLANE and Mr. H. G. HAY. Wave Propagation in a Slipping Stream of Electrons: Small Amplitude Theory	409
Dr. J. A. GLEDHILL and Dr. M. E. SZENDREI. Theory of the Production of an Ionized Layer in a Non-Isothermal Atmosphere Neglecting the Earth's Curvature, and its Application to Experimental Results	427
Mr. J. E. H. BRAYBON. A New Method of Measurement of the Variation with Wavelength of the Refractive Index and Absolute Stress Optical Coefficients of Amorphous Solids	446
Dr. MARY D. WALLER. Vibrations of Free Elliptical Plates	451
Letters to the Editor:	
Dr. D. K. C. MACDONALD and Mr. J. E. STANWORTH. Preparation of Alkali Metals in Glass	455
Mr. W. H. SHORTT. The Dissipation of Energy by a Pendulum Oscillating in Air at Low Pressures	456
Reviews of Books	458
Contents for Section A	461
Abstracts for Section A	462

Part 7. July 1950

Prof. G. I. FINCH. The Sliding Surface (34th Guthrie Lecture)	465
Mr. J. CRANK. The Influence of Concentration-Dependent Diffusion on Rate of Evaporation	484
Mr. C. R. BARBER. The E.M.F.-Temperature Calibration of Platinum, 10% Rhodium-Platinum and Platinum, 13% Rhodium-Platinum Thermocouples over the Temperature Range 0°-1,760° c.	492
Dr. J. ASHMEAD. A Joule-Thomson Cascade Liquefier for Helium	504
Dr. R. STREET and Mr. J. C. WOOLLEY. Time Decrease of Magnetic Permeability in Alnico	509
Prof. L. F. BATES and Mr. J. R. MALLARD. The Magnetic Properties of Uranium and Uranium-Iron Alloys	520
Dr. E. H. LINFOOT. On Phase-Contrast Testing with a Slit Source	527
Letters to the Editor:	
Mr. P. C. BANBURY and Dr. H. K. HENISCH. On the Frequency Response of PbS Transistors	540
Reviews of Books	541
Contents for Section A	542
Abstracts for Section A	542

Part 8. August 1950

Prof. S. TOLANSKY and Mr. N. BARAKAT. New Localized Multiple-Beam Interference Fringes formed with Curved Thin Sheets	545
Mr. P. A. WAYMAN. The Monocentric Schmidt-Cassegrain Cameras	553
Mr. A. E. J. HOLTHAM and Mr. H. A. PRIME. The Operation and Photographic Characteristics of a Kerr-cell Type of Electro-optical Shutter	561
Mr. F. J. BRADSHAW. The Optical Emissivity of Titanium and Zirconium	573
Dr. R. BECHMANN. Determination of the Elastic and Piezoelectric Coefficients of Monoclinic Crystals, with particular Reference to Ethylene Diamine Tartrate	577
Mr. N. W. RAMSEY. Some Measurements of the Resistivity of Good Insulators	590

	PAGE
Mr. B. MILNES and Mr. R. S. UNWIN. A Radio Meteorological Investigation in the South Island of New Zealand	595
Mr. C. C. NEWTON, Mr. F. J. HYDE and Mr. H. G. FOSTER. Ionospheric Cross-Modulation : Techniques of Measurement	616
Letters to the Editor :	
Dr. R. STREET and Mr. P. D. WHITAKER. The Measurement of Microwave Power at Wavelengths of 3 cm. and 10 cm.	623
Dr. E. SCHWARZ. Photoconductive Cells of Cadmium Selenide	624
Reviews of Books	626
Contents for Section A	629
Abstracts for Section A	630

Part 9. September 1950

Dr. J. B. HIGHAM and Prof. J. M. MEEK. Voltage Gradients in Long Gaseous Spark Channels	633
Dr. J. B. HIGHAM and Prof. J. M. MEEK. The Expansion of Gaseous Spark Channels	649
Dr. C. DODD. The Electrical Resistance of Liquid Gallium in the Neighbourhood of its Melting Point	662
Mr. B. COLLINGE. Hydrogen-filled Geiger Counters	665
Dr. B. E. NOLTINGK and Mr. E. A. NEPPIRAS. Cavitation produced by Ultrasonics.	674
Mr. H. G. W. HARDING. The Colour Temperature of Light Sources	685
Dr. P. A. LINDSAY. Certain Properties of Electrostatic Fields Encountered in Electron Lenses	699
Mr. U. F. GIANOLA. Reduction of the Spherical Aberration of Magnetic Electron Lenses	703
Mr. W. WEINSTEIN. The Computation of Wave-Front Aberrations of Oblique Pencils in a Symmetrical Optical System	709
Letters to the Editor :	
Mr. E. O. HALL. An Optical Method for Studying the Deformation of Mild Steel	724
Mr. S. N. DENNO, Mr. H. A. PRIME and Dr. J. D. CRAGGS. The Scattering of 3-cm. Radiation by Ionized Gases	726
Mr. W. R. LOOSEMORE and Dr. DENIS TAYLOR. Temperature Dependence of Counter Characteristics in Self-Quenching Geiger-Müller Counters	728
Reviews of Books	730
Contents for Section A	734
Abstracts for Section A	734

Part 10. October 1950

Dr. H. H. HOPKINS and Mr. P. M. BARHAM. The Influence of the Condenser on Microscopic Resolution	737
Dr. H. KUHN and Mr. B. A. WILSON. Reflectivity of Thin Silver Films and their Use in Interferometry	745
Dr. A. F. GIBSON. The Absorption Spectra of Solid Lead Sulphide, Selenide and Telluride	756
Dr. M. S. B. CHAGHTAI. Note on the Electron Velocity Distribution in Low Voltage Arcs	768
Mr. M. R. HOPKINS and Mr. T. C. TOYE. The Determination of the Viscosity of Molten Metals	773
Dr. R. MILLERSHIP and Mr. F. V. WEBSTER. High Frequency Permeability of Ferromagnetic Materials	783
Mr. K. S. W. CHAMPION. The Production of Pulsed Magnetic Fields using Condenser Energy Storage	795
Dr. G. G. MACFARLANE. A Theory of Contact Noise in Semiconductors	807
Dr. C. DODD and Mr. G. N. ROBERTS. Dielectric Loss and Dielectric Constant Measurements in Supercooled Liquids	814

Contents

vii
PAGE

Letters to the Editor :

Mr. P. D. LOMER. The Dielectric Strength of Aluminium Oxide Films	818
Mr. G. D. ADAM and Dr. K. J. STANDLEY. Ferromagnetic Resonance in Manganese Arsenide	820
Mr. J. T. KENDALL. Electrical Conductivity of Gray Tin	821
Mr. G. FRANCIS and Dr. A. VON ENGEL. Development of the Low Pressure Electrodeless Discharge in a High-Frequency Electric Field	823
Reviews of Books	824
Contents for Section A	830
Abstracts for Section A	831

Part 11. November 1950

Dr. H. R. THIRSK. A Note on the Orientated Overgrowths of Metal Films on Single Crystal Inorganic Substrates	833
Dr. D. J. PHILLIPS and Dr. N. THOMPSON. Surface Effects in Creep of Cadmium Crystals	839
Dr. M. DAVIS and Dr. N. THOMPSON. Creep in a Precipitation-Hardened Alloy	847
Dr. H. K. HENISCH and Mr. J. EWELS. A Study of Electrical Forming Phenomena at Selenium Contacts	861
Dr. D. GREENE. Secondary Electron Emission from Molybdenum produced by Helium, Neon, Argon and Hydrogen	876
Mr. A. C. LYNCH. The Variation with Temperature of the Piezoelectric Coefficients of Quartz	890
Mr. I. G. ROSS and Dr. R. A. SACK. Solvent Effects in Dipole Moment Measurements	893
Mr. B. H. BRIGGS and Mr. G. J. PHILLIPS. A Study of the Horizontal Irregularities of the Ionosphere	907
Dr. S. R. KHASTGIR and Mr. P. M. DAS. Periodic Fading of Short-Wave Radio Signals	924
Dr. J. MCG. BRUCKSHAW and Dr. B. S. RAO. Magnetic Hysteresis of Igneous Rocks	931
Mr. W. CULSHAW. The Michelson Interferometer at Millimetre Wavelengths	939
Mr. P. A. STURROCK. Note on the Focusing of Electron Beams in certain Magnetic Fields	954
Dr. J. F. W. BELL. Satellite Resonances in Ultrasonic Interferometry	958
Letters to the Editor :	
Dr. J. R. BRISTOW: Dr. R. C. PARKER and Mr. D. HATCH. Frictional Relaxation Oscillations	964
Contents for Section A	965
Abstracts for Section A	966

Part 12. December 1950

Dr. P. JACQUINOT. Quelques recherches sur les raies faibles dans les spectres optiques (Fifth Holweck Discourse)	969
Dr. M. R. A. RAO. Electron Microscopic Studies on Aqueous Sols	980
Mr. T. S. MOSS. Photoconductivity in Magnesium Antimonide Layers	982
Prof. E. N. DA C. ANDRADE, Dr. R. F. Y. RANDALL and Mr. M. J. MAKIN. The Reh binder Effect	990
Mr. R. PARKER. A Study of the Magneto-Resistance of Silicon-Iron	996
Dr. R. S. TEBBLE and Dr. W. D. CORNER. Investigations on the Reversible Susceptibility of Ferromagnetics	1005
Dr. G. B. WALKER. Congruent Space Charge Flow	1017
Mr. J. M. DANIELS. A 100-Kilowatt Water-cooled Solenoid	1028
Letters to the Editor :	
Mr. H. D. KEITH. Precision Lattice-Parameter Measurements	1034
Mr. R. E. BURGESS. The Capacitance of Selenium Rectifiers	1036

	PAG
Mr. U. F. GIANOLA. The Correction of the Spherical Aberration of Electron Lenses using a Correcting Foil Element	103
Mr. J. PATON. Aurora and Luminous Night Clouds	103
Mr. F. V. WEBSTER and Mr. K. S. DRIVER. High Frequency Permeability of Ferromagnetic Materials	104
Obituary Notices :	
JAMES REGINALD ASHWORTH	104
ALFRED JOSEPH BULL	104
FREDERICK CHARLES CLARKE	104
JAMES ARNOLD CROWTHER	104
GEOFFREY E. F. FERTEL	104
CYRIL OWEN GREEN	104
HERMAN SHAW	104
GEORGE WILLIAM TODD	104
ALEXANDER WOOD	104
Contents for Section A	105
Abstracts for Section A	105
Subject Index, Section B, Vol. 63	105
Index of Authors (with Titles), Section B, Vol. 63	105
Index to Reviews of Books, Section B, Vol. 63	106

THE PROCEEDINGS OF THE PHYSICAL SOCIETY

Section B

VOL. 63, PART 1

1 January 1950

No. 361 B

EDITORIAL

PROCEEDINGS OF THE PHYSICAL SOCIETY SECTIONS A and B

In the light of experience gained over the course of the past year it is now possible to make a clearer statement of the basis on which division into the two Sections of the *Proceedings* is being made.

In general, it has been found that the division which best meets the needs of Fellows is one which brings papers on microphysics and the physics of elementary particles into the one Section, A, and papers on macroscopic physics into the other Section, B. While it is essential to maintain a certain amount of flexibility in the allocation of papers, a paper will, as far as possible, be put in the Section where the main interest lies, e.g. a paper on counter-technique would appear in B if the circuitry is the main interest, but would be in A if the applications to actual counting are of major importance.

A list indicating this subject division is given below.

A. C. STICKLAND,
Secretary-Editor.

Section A

Thermodynamics
Electrodynamics
Statistical mechanics
Quantum theory
Nuclear physics
Electron theory: metals, semiconductors, dielectrics
Molecular and atomic structure
Spectra of atoms and molecules over the whole frequency range

Physics of crystals (including luminescence as a means of investigating this)
Photoconductivity
Theory of solids, liquids and gases
Magnetism (theoretical)
Cosmic rays
Low temperature physics
Standards relevant to this Section

Section B

Thermal, magnetic, electrical, elastic, rheological and other mechanical properties of matter
Hydrodynamics and aerodynamics
Acoustics
Metallography
Geometrical optics, optical design and microscopy
Interferometry, diffraction and classical physical optics generally
Electric and electronic circuits
Dielectrics and semiconductors: measurement and theory of measurement
Electric discharges
Electron optics
Geophysics
Radio
Physics of the atmosphere and ionosphere
Physics of the sensory processes

Colour physics
Astrophysics and solar physics
Applications of spectroscopy (e.g. chemical analysis, structure analysis)
Analysis of crystal structure
Radiography
Luminescence
Photoelectricity
Thermoelectricity
Piezoelectricity
Photoconductivity
Magnetic properties of materials (including methods of measurement)
Production of low temperatures
High pressure physics
Vacuum physics
Thermionic emission
Standards relevant to this Section

The Elastic Constants of a Solid containing Spherical Holes

By J. K. MACKENZIE *

H. H. Wills Physical Laboratory, University of Bristol

MS. received 4th August 1949

ABSTRACT. The effective bulk and shear moduli are calculated by a self-consistent method due to Fröhlich and Sack. The bulk modulus k is determined by applying a hydrostatic pressure, and the shear modulus μ by applying a simple homogeneous shear stress, to a large sphere. Each hole is surrounded by a spherical shell of real material, and the reaction of the rest of the material is estimated by replacing it by equivalent homogeneous material. For consistency, both the density and the displacement of the outer spherical boundary must be the same whether the hole and its surrounding shell are replaced by equivalent material or not. The effective elastic constants calculated from these conditions are

$$\begin{aligned} 1/k &= 1/k_0\rho + 3(1-\rho)/4\mu_0\rho + O[(1-\rho)^3], \\ (\mu_0 - \mu)/\mu_0 &= 5(1-\rho)(3k_0 + 4\mu_0)/(9k_0 + 8\mu_0) + O[(1-\rho)^2], \end{aligned}$$

where k_0 and μ_0 refer to the real material and ρ is the density of the actual material relative to that of the real material; in the next approximation k depends on the standard deviation of the volumes of the holes.

The dilatation due to a distribution of pressures in the holes is $\bar{p}(1/k - 1/k_0)$, where \bar{p} is the mean obtained when the pressure in each hole has a weight proportional to the volume of the hole. By using the hydrodynamic analogue of the elastic problem, the theory is briefly applied to the theory of sintering, and used to discuss the effective viscosity of a liquid containing small air bubbles.

§ 1. INTRODUCTION

WHILE developing a theory of sintering (Mackenzie and Shuttleworth 1949) it became necessary to calculate the compressibility of a solid containing spherical holes. In this paper all the holes were assumed to be of the same size and embedded in an incompressible medium. In this simple case it suffices to equate the stored elastic energy to the work done by the externally applied pressure in order to derive the effective compressibility. The original calculation, however, followed the lines of ordinary elasticity theory, and both the effective compressibility and the effective shear modulus were derived without assuming either holes of equal size or incompressibility of the real material. It is these calculations which are reported in the present paper. On the other hand, the present method can only be used when the relation between stress and strain is linear, while the energy method used in the paper on sintering is applicable when this relation is non-linear.

In §2 the problem is analysed in some detail and the self-consistent method of calculation is described. In §3 two problems are considered: (a) the calculation of the compressibility of a material full of small spherical holes of different sizes, and (b) the calculation of the dilatation due to different pressures acting in the holes; the shear modulus of a material full of small spherical holes is calculated in §4. Finally, in §5 the close similarity between the equations of hydrodynamics and elasticity is used to discuss briefly an application to the theory of sintering and to discuss the effective viscosity of a liquid containing small air bubbles.

The results of these calculations show that, to a good approximation, both the effective elastic moduli depend only on the relative density ρ and the elastic

* Now at C.S.I.R.O., Division of Tribophysics, Melbourne, Australia.

constants of the real material; holes have a comparatively large effect on compressibility and only a small effect on the shear modulus. It may be conjectured that this is true in general no matter what the shape or size of the individual holes may be, provided of course that extremes are excluded. It is not clear how the numerical constants appearing in the results will depend on the particular shape of hole chosen for calculation, but it seems likely that the formulae obtained here will give the correct order of magnitude in most cases.

§ 2. THE METHOD OF CALCULATION

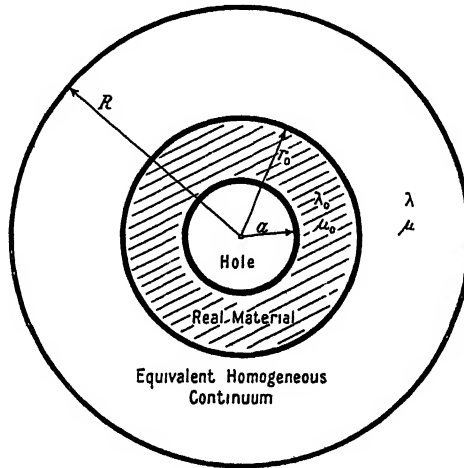
It will be assumed that the real material contains isolated spherical holes distributed at random throughout the volume of the material, and that the real material has homogeneous and isotropic elastic properties. It will also be assumed that the volume of all the holes is small compared with the total volume, and that the total volume contains a large number of holes. The calculation uses the notion of an equivalent homogeneous continuum introduced in the next paragraph.

If attention is directed not to the detailed structure of the actual solid but to a volume element containing a large number of holes, then, because the distribution of the size of holes and of their position in space is statistically uniform, such a volume element can be regarded as forming part of a homogeneous isotropic continuum. The elastic constants of this equivalent homogeneous continuum will be calculated in terms of the elastic properties of the real material and the number and size distribution of the holes. When large volumes are concerned, the elastic properties of the actual solid will be independent of its external shape, and for simplicity it is convenient to choose this shape to be spherical. The effective elastic constants will be calculated by applying a stress to the outer spherical boundary and comparing the strains produced in the actual solid with the strains produced by the same stress in a homogeneous isotropic solid. A uniform hydrostatic pressure will be applied to determine the effective bulk modulus and a simple homogeneous shear stress (with no hydrostatic component) to calculate the effective shear modulus. The results obtained depend to some extent on the particular stress which is applied and, in general, the more complicated the stress the more the detailed structure of the actual solid will become apparent.

A convenient self-consistent method of calculation has been indicated by Fröhlich and Sack (1946). Their method is equivalent to a perturbation calculation, and is certainly valid when the volume of all the holes is sufficiently small compared with the total volume. The method depends on the fact that any macroscopic volume (i.e. one containing a large number of holes) can be replaced by equivalent homogeneous material without essentially changing the elastic behaviour of the solid. This is so because the mean stresses and displacements at the boundary of the volume are equal to those at the boundary of the same volume in the equivalent elastic continuum. The conditions for consistency, namely that the density and the displacement of the outer boundary shall be unchanged by such replacement, enable the effective elastic constants to be calculated. The particular method of calculation adopted is described in detail in the next paragraph.

Since the holes are isolated, every hole of radius a will be surrounded by a spherical shell of real material out to at least some radius r_0 (see Figure), and this may be regarded as a typical volume element of the actual solid. The model used for the calculation of the effective elastic constants is then as follows. A hole of radius a is surrounded by a spherical shell of real material out to radius r_0 , and this

in turn is surrounded by a spherical shell of equivalent homogeneous material (whose elastic constants are as yet unknown) out to some large radius R . A particular stress is applied to this outer spherical boundary and the displacements of the outer surface calculated; this displacement is then weighted according to the frequency of occurrence of holes of radius a and averaged over the size distribution of the holes. Finally, the resulting mean displacement is equated to the displacement which would have resulted had the sphere of radius R been completely filled with equivalent homogeneous material. Thus, one of the consistency conditions is satisfied and an equation for the effective elastic constants derived.



The model used for calculation.

The other condition determines r_0 , since the density of a hole and its surrounding shell must be the same as that of the actual material. If $f(a)da$ is the proportion of holes with radii lying in the range $(a, a + da)$ then the volume of holes per unit volume of the real material is

$$\frac{1-\rho}{\rho} = \frac{4}{3}\pi n \int f(a)a^3 da = \frac{1}{\rho r_0^3} \int f(a)a^3 da, \quad \dots\dots(1)$$

where the integration is over the whole of the relevant range of a , ρ is the relative density, and n is the number of holes per unit volume of the real material. Note that $\int f(a) da = 1$.

In general, when part of a solid is replaced by material of different elastic constants there is an additional term in the displacement of the outer boundary of the order of the ratio of the volume replaced to the total volume. If, however, the elastic constants of the new material are the same as the effective constants of the original material this term will disappear. In fact, the consistency condition will be expressed by equating this term of order r_0^3/R^3 to zero. In addition, there will usually be terms of higher order which arise because the boundary conditions are never met exactly when part of the actual solid is replaced by equivalent continuum, but only on the average. These terms will be neglected.

In the following sections the method outlined above will be applied to the calculation of the bulk and shear moduli of a material containing small spherical holes.

§ 3. THE COMPRESSIBILITY

Two problems will be considered in this section: (a) the calculation of the compressibility of a material containing small spherical holes of different sizes, and (b) the calculation of the dilatation due to a distribution of pressures inside the holes; these pressures may be due to included gas or to surface tension in the boundary of the hole. The compressibility will be calculated by applying a hydrostatic pressure P to the outer spherical boundary and using the method described in the last section. The same model is also applicable when a pressure p is applied inside the central hole with zero pressure in all the other holes; in this case, the approximation consists in estimating the reaction of the part of the actual solid which lies outside a sphere of radius r_0 by replacing it by equivalent homogeneous material. When there are pressures in all the holes the total displacement is found by adding the displacements due to each hole individually.

The detailed calculations fall naturally into three parts. In § 3 (i) the general equations for the elastic displacements will be written down, while in the two following sub-sections the effect of an externally and an internally applied pressure will be considered in turn. However, before proceeding it is convenient to show how the second problem may be reduced to the first.

Consider first the case where the pressure p is the same in every hole and there is no externally applied pressure. If an additional hydrostatic pressure $-p$ is applied to the surfaces of all the holes and to the external surface, then the resultant pressure is zero in all the holes and $-p$ on the external surface. Now when a hydrostatic pressure is applied to all the boundaries of a homogeneous elastic material the stress at all points inside the material is a hydrostatic pressure equal to that applied (Love 1944, p. 86). Hence, the displacements due to the extra hydrostatic pressure $-p$ can readily be estimated by filling up all the holes with real material and applying the pressure $-p$ to the outer boundary. (A comparison of the two situations shows that all the boundary conditions are met.) Thus the total dilatation is $p(1/k - 1/k_0)$, and the additional term is zero if the material is incompressible.

On the other hand, if the pressure is not the same in all the holes, then a fixed part of this pressure can be transferred to the outer boundary as explained in the last paragraph, and the effect of the pressures remaining in the holes estimated separately. If these remaining pressures produce no total dilatation they may be neglected. Now the displacement of the outer boundary when only one particular hole has a pressure in it is proportional to the pressure in the hole and to its volume. Therefore, if the pressure in each hole is given a weight proportional to the volume of the hole, then when the mean pressure so calculated is transferred to the outside the remaining distribution of pressures will produce no displacement of the external boundary. This statement will be verified in § 3 (iii).

(i) *General Elastic Equations*

First a word about notation. Symbols with subscript zero refer to the real material or to the intermediate boundary, while the same symbol without a subscript usually denotes the corresponding magnitude for the equivalent homogeneous continuum. Thus λ and μ are the effective elastic constants which are to be calculated, while λ_0 and μ_0 are the same constants for the real material; the constant μ is the rigidity and the bulk modulus $k = \lambda + \frac{2}{3}\mu$.

It is required to find the displacements in the composite body illustrated in the Figure when hydrostatic pressures are applied to both boundaries. On account of the spherical symmetry, the displacements will be purely radial and given by (Love 1944, p. 142)

$$u = Ar + B/r^2,$$

while the normal pressure acting across the spherical boundary at the same radius is

$$(\lambda + 2\mu) \frac{\partial u}{\partial r} + 2\lambda \frac{u}{r}.$$

Then, using Love's results for a thick spherical shell, the radial displacement in the outer shell of equivalent homogeneous material is

$$u = \frac{r}{3k} \frac{p_0 r_0^3 - PR^3}{R^3 - r_0^3} + \frac{R^3 r_0^3}{4\mu r^2} \frac{p_0 - P}{R^3 - r_0^3},$$

and the radial displacement in the inner shell of real material is

$$u = \frac{r}{3k_0} \frac{pa^3 - p_0 r_0^3}{r_0^3 - a^3} + \frac{a^3 r_0^3}{4\mu_0 r^2} \frac{p - p_0}{r_0^3 - a^3},$$

where p is the pressure inside the hole (all other holes at zero pressure) and P is the externally applied pressure.

The pressure p_0 at the intermediate boundary is determined by the boundary condition that requires the radial displacement to be continuous across this boundary. This condition gives, neglecting r_0^3/R^3 and higher powers,

$$p_0 \left[\frac{1}{4\mu} + \frac{1}{3k_0} \frac{r_0^3}{r_0^3 - a^3} + \frac{1}{4\mu_0} \frac{a^3}{r_0^3 - a^3} \right] = P \left[\frac{1}{3k} + \frac{1}{4\mu} \right] + p \left[\frac{1}{3k_0} + \frac{1}{4\mu_0} \right] \frac{a^3}{r_0^3 - a^3}. \quad \dots (2)$$

This equation for p_0 , together with the expression for the displacement of the external boundary, namely

$$\frac{u}{R} = -\frac{P}{3k} + (p_0 - P) \left[\frac{1}{3k} + \frac{1}{4\mu} \right] \frac{r_0^3}{R^3 - r_0^3}, \quad \dots (3)$$

completely determines the external change of volume. The detailed solution of the two problems can now be obtained quite simply.

(ii) Pressure applied Externally

In this sub-section the effective compressibility k is determined when there are no pressures in the holes. When a pressure P is applied to the external surface, the resulting displacement of the external boundary due to holes of some particular radius a is given by equation (3). Averaging this displacement over the size distribution of the holes and equating the resulting mean displacement to the displacement of the external boundary when a pressure P is applied to a solid sphere of the equivalent homogeneous material, viz. $-PR/3k$, as required by the consistency condition, it follows that

$$P = \int p_0 f(a) da, \quad \dots (4)$$

where p_0 is given by equation (2) and $\int f(a) da = 1$. If the holes are all the same size, $p_0 = P$, and (7) (with $\sigma = 0$) follows from (2) immediately.

On putting $p = 0$ into equation (2) and expanding in powers of a^3/r_0^3 ,

$$\frac{p_0}{P} = \frac{1 + 4\mu/3k}{1 + 4\mu/3k_0} \left[1 - \frac{a^3}{r_0^3} + \frac{(\mu_0 - \mu)/\mu_0}{1 + 4\mu/3k_0} \left\{ \frac{a^3}{r_0^3} - \frac{a^6}{r_0^6} \right\} + \dots \right].$$

Now it will be shown in § 4 that

$$(\mu_0 - \mu)/\mu_0 = \kappa(1 - \rho), \quad \dots\dots(5)$$

where κ depends only on k_0 and μ_0 and is given by equation (20). Put

$$\int \frac{a^3}{r_0^3} f(a) da = 1 - \rho \quad \dots\dots(1)$$

and

$$\int \frac{a^6}{r_0^6} f(a) da = \sigma^2 + (1 - \rho)^2, \quad \dots\dots(6)$$

where $4\pi\sigma r_0^3/3$ is the standard deviation of the volumes of the holes. Then it follows from (4), after some reduction, that

$$\frac{1}{k} = \frac{1}{k_0\rho} + \frac{3}{4\mu_0} \frac{1 - \rho}{\rho} (1 + \kappa\sigma^2). \quad \dots\dots(7)$$

This equation gives the effective compressibility up to and including terms of order $(1 - \rho)^3$. The value of κ is always about 2. It should be noted that a few holes in an otherwise incompressible medium can give rise to an appreciable finite compressibility.

It is not altogether certain that the model is adequate for the calculation of a third approximation to k . It is clear, however, that the effect of a distribution in the sizes of the holes is small, and that to a good approximation the effective compressibility depends only on ρ and the elastic constants of the real material. In fact, (7) gives the correct behaviour for the effective compressibility in the extreme cases as the density tends to unity or to zero. Thus it is plausible that this equation should be roughly correct for all relative densities.

(iii) Pressure applied Internally

In this sub-section the effect of a distribution of pressures inside the holes is considered. If there are different pressures in holes of the same size it is clear that their effect is the same as the mean pressure \bar{p} acting in all of them. When a pressure p is applied in only one hole, equation (3), with $P = 0$, gives the resulting displacement of the external boundary. Then, on multiplying by the number of holes, R^3/r_0^3 , and averaging over the size distribution, the mean displacement of the external surface is (neglecting r_0^3/R^3 compared with unity)

$$u = R \left[\frac{1}{3k} + \frac{1}{4\mu} \right] \int p_0 f(a) da = \bar{p} R \left[\frac{1}{3k} - \frac{1}{3k_0} \right], \quad \dots\dots(8)$$

where \bar{p} is the mean pressure which is transferred to the external surface, and the last part of the equation follows from the discussion at the end of the introductory part of § 3. The last equation in (8) determines the mean effective pressure \bar{p} , which will now be calculated.

On putting $P=0$ into equation (2), substituting the value of p_0 into (8), and expanding in powers of a^3/r_0^3 , it follows that

$$\bar{p} = \frac{1/3k_0 + 1/4\mu_0}{1/3k - 1/3k_0} \frac{1 + 4\mu/3k}{1 + 4\mu/3k_0} \int p \frac{a^3}{r_0^3} \left[1 + \frac{(\mu_0 - \mu)/\mu_0}{1 + 4\mu/3k_0} \frac{a^3}{r_0^3} \right] f(a) da.$$

Then, substituting for k from (7) and replacing p by \bar{p} when evaluating the correction term in the integral, it follows, after some reduction, that

$$\bar{p} = \frac{1}{a_0^3} \int p a^3 f(a) da, \quad \dots\dots(9)$$

where $a_0^3 = (1 - \rho)r_0^3$, or $4\pi a_0^3/3$ is the mean volume of the holes. Although this expression is correct up to and including terms of order $(1 - \rho)^3$, it is not clear what error has been committed in replacing p by \bar{p} in the small correction term. Equation (9) justifies the statement made at the very end of the introductory part of § 3.

Finally, returning to equation (8), it is clear that the dilatation caused by a distribution of pressures in the holes is \bar{p}/k' , where \bar{p} is given by (9) and

$$\frac{1}{k'} = \frac{1}{k} - \frac{1}{k_0} = \frac{1 - \rho}{\rho} \left[\frac{1}{k_0} + \frac{3}{4\mu_0} (1 + \kappa\sigma^2) \right]. \quad \dots\dots(10)$$

§ 4. THE SHEAR MODULUS

The effective shear modulus will now be calculated by applying a simple type of homogeneous shear stress and calculating the resulting displacements of the external boundary both for the composite solid shown in the Figure and for a sphere of equivalent homogeneous material. As before, comparison of the two solutions gives an equation from which the effective shear modulus can be calculated.

Now the displacements which result from all the simple types of homogeneous shearing stress can be derived from solid spherical harmonics of degree two. Such a solid harmonic is of the form $r^2 S_2$, where S_2 is a surface harmonic of degree two and determines the symmetry of the displacements; the solid harmonic $r^{-3} S_2$ also has the same symmetry, and will therefore be expected to occur in the general solution.

The solution of the elastic equations, which is derived from the solid harmonic $r^2 S_2$, and for which the dilatation and the rotation are everywhere zero, is (Love 1944, p. 250)

$$\mathbf{u} = \text{grad}(r^2 S_2) = 2r S_2 \mathbf{r} + r S_2' \boldsymbol{\tau}, \quad \dots\dots(11)$$

and the corresponding stress across a spherical boundary at the same radius is

$$\mathbf{F} = 2\mu r^{-1} \text{grad}(r^2 S_2) = 4\mu S_2 \mathbf{r} + 2\mu S_2' \boldsymbol{\tau}, \quad \dots\dots(12)$$

where \mathbf{r} and $\boldsymbol{\tau}$ are unit vectors normal and tangential to the spherical boundary and S_2' is independent of r . This solution gives the displacements when a shearing stress \mathbf{F} , which is the same at all points, is applied to a solid homogeneous sphere. This stress will now be applied to the composite solid and the resulting displacements calculated.

There are four independent solutions of the elastic equations which can be derived from solid harmonics with the same symmetry as S_2 , and for which the rotation is everywhere zero. Writing

$$p S_2 \mathbf{r} + q S_2' \boldsymbol{\tau} = \{p, q\}, \quad \dots\dots(13)$$

these solutions are (Love 1944, p. 250):

$$\left. \begin{aligned} \mathbf{u}_1 &= \text{grad}(r^2 S_2) &= \{2, 1\}r, \\ \mathbf{u}_2 &= \text{grad}(r^{-3} S_2) &= \{-3, 1\}/r^4, \\ \mathbf{u}_3 &= r^2 \text{grad}(r^2 S_2) + \alpha_2 r^3 S_2 \mathbf{r} &= \left\{ \frac{6\lambda}{5\lambda + 7\mu}, 1 \right\} r^3, \\ \mathbf{u}_4 &= r^2 \text{grad}(r^{-3} S_2) + \alpha_{-3} r^{-2} S_2 \mathbf{r} &= \left\{ \frac{3\lambda + 5\mu}{\mu}, 1 \right\} / r^2; \end{aligned} \right\} \dots\dots (14)$$

the corresponding stresses acting across a spherical surface of radius r are, respectively,

$$\left. \begin{aligned} \mathbf{F}_1/\mu &= \{4, 2\}, \\ \mathbf{F}_2/\mu &= \{24, -8\}/r^5, \\ \mathbf{F}_3/\mu &= \left\{ \frac{-6\lambda}{5\lambda + 7\mu}, \frac{16\lambda + 14\mu}{5\lambda + 7\mu} \right\} r^2, \\ \mathbf{F}_4/\mu &= \left\{ -\frac{18\lambda + 20\mu}{\mu}, \frac{3\lambda + 2\mu}{\mu} \right\} / r^3. \end{aligned} \right\} \dots\dots (15)$$

Thus the general solution of the elastic equations in a thick spherical shell will be of the form

$$\mathbf{u} = A\mathbf{u}_1 + B\mathbf{u}_2 + C\mathbf{u}_3 + D\mathbf{u}_4. \quad \dots\dots (16)$$

This solution will apply as it stands in the outer shell containing equivalent homogeneous material; the corresponding solution for the inner shell containing the real material is derived simply by putting a subscript zero on all the constants. The whole solution then involves eight arbitrary constants which are determined by eight equations arising from the boundary conditions; two equations have to be satisfied for each of four boundary conditions.

The four boundary conditions are: (a) that the stress across the external boundary is $\mu\{4, 2\}$, (b) that the displacements are continuous across the intermediate boundary, (c) that the stress is continuous across the intermediate boundary, and (d) that the innermost boundary is stress-free. Using equations (14), (15) and (16), these boundary conditions lead to a set of equations whose formal solution can be written down immediately in a form involving eighth-order determinants. The solution can then be expanded quite simply in powers of r_0/R , and the result of doing this is

$$\left. \begin{aligned} A - 1 &= K_1 \frac{\Delta_2}{\Delta_1} \frac{r_0^3}{R^3}, & B/R^5 &= 0, \\ CR^2 &= K_2 \frac{\Delta_2}{\Delta_1} \frac{r_0^3}{R^3}, & D &= \frac{\Delta_2}{\Delta_1} \frac{r_0^3}{R^3}, \end{aligned} \right\} \dots\dots (17)$$

where powers of r_0/R higher than the third have been neglected, Δ_1 and Δ_2 are certain determinants of the sixth order, and K_1, K_2 involve only λ and μ . It is clear from this solution that the additional terms in the displacement due to the presence of the hole and its surrounding shell are of order r_0^3/R^3 , as was stated at the end of §2.

Substituting the values of the constants into (16), integrating over the size distribution, and comparing the results with the displacement given by equation (11), it follows that the condition for consistency is

$$\int \frac{\Delta_2}{\Delta_1} f(a) da = 0. \quad \dots\dots (18)$$

Finally, expanding Δ_1 and Δ_2 in powers of a/r_0 , and neglecting powers higher than the third in the quotient Δ_2/Δ_1 , it follows that

$$\frac{\mu_0 - \mu}{\mu_0} = 5 \frac{3k_0 + 4\mu_0}{9k_0 + 8\mu_0} (1 - \rho). \quad \dots\dots (19)$$

This equation gives the fractional decrease in the modulus of rigidity due to the presence of the holes. The value of the constant κ introduced in equation (5) is

$$\kappa = 5 \frac{3k_0 + 4\mu_0}{9k_0 + 8\mu_0}. \quad \dots\dots (20)$$

§ 5. APPLICATIONS

The application of the results of the preceding calculations to determining the effective elastic constants of materials containing a large number of small holes is obvious and will not be discussed further. There is, however, a close similarity between the equations of elasticity and those of hydrodynamics (Goodier 1936), and it is the results which can be deduced from this similarity which will be briefly discussed in the following paragraphs.

If, in the hydrodynamic equations for slow steady motion (so that $D/Dt = 0$), the velocity \mathbf{v} is replaced by the displacement \mathbf{u} , the viscosity η by the rigidity μ , and $-\mathbf{p}$ by $k \operatorname{div} \mathbf{u}$, then the equations of elasticity are obtained, and vice versa. Thus, by means of the above scheme of replacement, the velocities in a slow and steady viscous motion due to some system of stresses can be derived immediately from the elastic displacements resulting from the same stresses in the corresponding elastic problem. When making this transfer the bulk modulus will usually be taken as infinite because, although the shearing viscosity of a liquid is finite, the 'volume viscosity' (which is the analogue of k) is usually considered to be infinite, i.e. there is no energy dissipated during a pure compression or expansion.

In the theory of sintering mentioned in the introduction, it was necessary to calculate the rate of increase of density of a compact consisting of spherical holes in an incompressible medium when a negative pressure $-2\gamma/a$ due to surface tension acts in all the holes. In the case of glass the action of these pressures is resisted by viscous forces, and when this is so the resulting flow may be derived from the results of § 3 (iii) by means of the above substitution. Thus the rate of increase of density is

$$\frac{d\rho}{dt} = - \frac{\rho \bar{p}}{k'}, \quad \dots\dots (21)$$

where \bar{p} is given by equation (9) as

$$\left. \begin{aligned} \bar{p} &= - \frac{2\gamma}{a_0^3} \int a^2 f(a) da \\ &= - \frac{2}{3} \gamma \frac{\text{mean surface area of holes}}{\text{mean volume of holes}} \end{aligned} \right\} \quad \dots\dots (22)$$

If all the holes are of the same size, then using (10) with $k_0 = \infty$ to determine k' ,

$$\frac{d\rho}{dt} = \frac{3}{2} \left(\frac{4\pi}{3} \right)^{\frac{1}{3}} \frac{\gamma n^{\frac{1}{3}}}{\eta} (1-\rho)^{\frac{2}{3}} \rho^{\frac{1}{3}}, \quad \dots\dots (23)$$

a formula previously derived in another way; n is the number of holes per unit volume of the real material.

Again, consider a liquid which contains a large number of small bubbles containing gas. These bubbles will be of such a size that the gas pressure is equal to the hydrostatic pressure in the liquid plus the pressure $2\gamma/a$ due to surface-tension. Now provided the liquid is sheared slowly and the bubbles retain approximately their spherical form, the analogue of equation (19) gives the fractional decrease in the viscosity due to the presence of the bubbles as

$$\frac{\eta_0 - \eta}{\eta_0} = \frac{5}{3} \alpha, \quad \dots\dots (24)$$

where α is the fraction of the total volume which is occupied by the bubbles. This expression assumes that the bubbles deform freely during the motion so that, unless the motion is oscillatory, the bubbles will not remain spherical for long. In practice, both the gas pressure and the surface tension will tend to maintain the shape of the bubble approximately spherical and equation (24) will not apply. Taylor (1932) has shown that the presence of liquid spheres which remain spherical during the motion raises the effective viscosity rather than lowers it. If the viscosity of the liquid in the spheres is η' , Taylor shows that

$$-\frac{\eta_0 - \eta}{\eta_0} = \frac{5}{2} \frac{\eta' + 2\eta_0/5}{\eta' + \eta_0} \alpha. \quad \dots\dots (25)$$

When the spheres are rigid, $\eta' = \infty$, and this reduces to Einstein's result (1906, 1911), and when $\eta' = 0$,

$$-\frac{\eta_0 - \eta}{\eta_0} = \alpha,$$

a result also obtained by Eizenschitz (1933) for the case where the liquid slips freely over the surface of a rigid sphere. It is clear therefore that boundary conditions assumed at the surface of the bubble play an important part in the results which are obtained.

ACKNOWLEDGMENT

The author wishes to thank Dr. R. Shuttleworth for a great deal of helpful discussion during the preparation of this paper.

REFERENCES

- EINSTEIN, A., 1906, *Ann. Phys., Lpz.*, **19**, 289 ; 1911, *Ibid.*, **34**, 591.
 EISENSCHITZ, R., 1933, *Phys. Z.*, **34**, 411.
 FRÖHLICH, H., and SACK, R., 1946, *Proc. Roy. Soc. A*, **185**, 415.
 GOODIER, J. N., 1936, *Phil. Mag.*, **22**, 678.
 LOVE, A. E. H., 1944, *Theory of Elasticity*, 4th edn. (New York : Dover).
 MACKENZIE, J. K., and SHUTTLEWORTH, R., 1949, *Proc. Phys. Soc. B*, **62**, 833.
 TAYLOR, G. I., 1932, *Proc. Roy. Soc. A*, **138**, 41.

On Movements of Small Ferromagnetic Particles in Inhomogeneous Magnetic Fields

By FRIEDRICH BLAHA

Institute of Physics, University of Vienna

Communicated by T. G. Hodgkinson ; MS. received 30th June 1949

ABSTRACT. Ferromagnetic particles of diameter from 10^{-3} to 10^{-5} cm. when exposed to light move along the lines of force of an inhomogeneous magnetic field. It does not seem possible to explain the movement when only dipole properties of the particles are taken into account. Field strengths of 10^{-2} gauss are sufficient to cause distinct movement. The motions resemble those of electrically charged particles in electric fields.

§ 1. INTRODUCTION

S MALL particles of iron powder show, when exposed to light in homogeneous magnetic fields, polar motions which are reversible with reversal of the field (magnetophotophoresis (Ehrenhaft 1930)). Since then many kinds of various substances have been found to show this behaviour. In continuation of researches on the phenomenon the following observations were carried out in simply formed, typical inhomogeneous fields. The method described for producing the particles is one of the most convenient and effective.

§ 2. EXPERIMENT

An electric arc between two iron electrodes in oxygen or air produces a sort of smoke consisting mainly of small particles of ferric oxide. Such particles of 10^{-3} to 10^{-5} cm. in diameter were collected in a small observation chamber which was surrounded by a small rectangular loop (3 mm. \times 10 mm.) of thin copper wire (see Figure 1). The chamber was specially constructed to allow of illumination and observation near the wire. The particles were observed in a dark field by means of a microscope (magnification 30 to 100 times), the axis of which was parallel to the plane of the above-mentioned loop. The assembly consisting of chamber, microscope and surroundings was iron-free and the geomagnetic field was compensated over the region of observation.

§ 3. OBSERVATIONS

Under the conditions of the experiment most of the particles appear as diffraction discs, others have the form of short, often crooked, threads. They all fall uniformly in the resisting medium (air) with a Brownian movement superimposed. In very few cases is common photophoresis (positive or negative) visible.

As soon as a direct current flows through the wire of the loop (for instance 0.2 amp.), forming an inhomogeneous magnetic field in the cell, 10 to 20% of the particles visible as discs start moving along the magnetic lines of force, some in, some against, the direction of the magnetic field. They pass the plane of the loop unhindered. When the direction of the current is reversed, these particles instantaneously reverse their direction of motion. They stop immediately the current is cut off. The other particles, so far as they appear as discs, do not show any reaction to the weak magnetic field applied.

All the particles shaped like short threads behave like tiny compass needles: pointing in the direction of the acting magnetic field they turn through 180° when the field is reversed. Among these particles also, some show the above described phenomenon, that is to say they move along the magnetic lines of force



Figure 3. Time of exposure 10 sec.
Magnification 40 times.

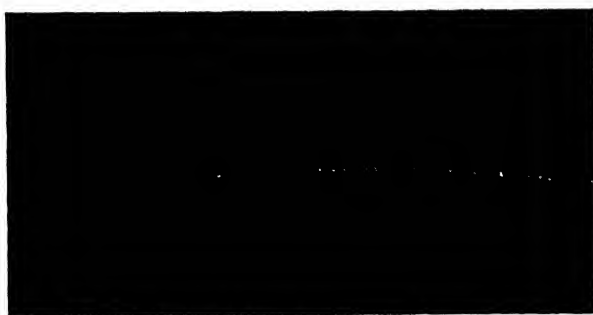


Figure 4. Time of exposure 2 sec.
Total magnification 120 times.



Figure 5. Time of exposure 0.4 sec.
Magnification 80 times. Sunlight.



Figure 6. Time of exposure 13 sec.
Magnification 80 times.

and pass the plane of the loop unhindered; reversal of the magnetic field causes a turn through 180° and reversal of the translatory movement.

Taking into account only dipole properties, from a theoretical point of view the particles should move from the right and the left towards the plane of the loop and the wire (provided that the gradient of the field is sufficient) without reversing their translatory movement.

The apparatus as above described did not permit the use of higher values of the inhomogeneous magnetic field. For this purpose the loop was replaced by a coil with iron core adjusted outside of the cell as shown in Figure 2. Small field strength (for example the remanent magnetism of the iron core) then caused movements analogous to those already described.

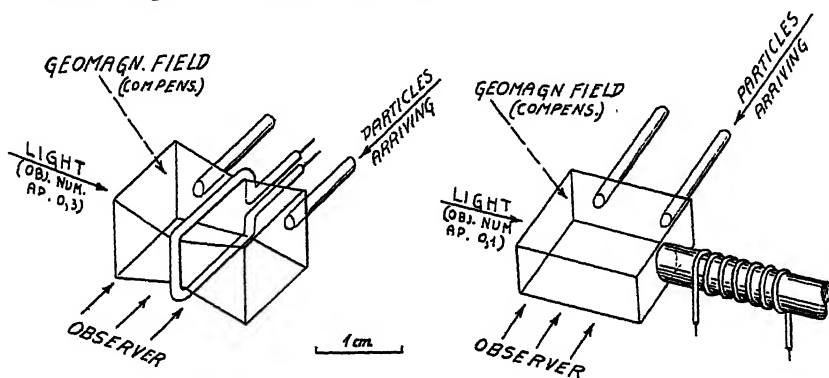


Figure 1.

Figure 2.

When now the magnetic field strength was gradually increased, the different kinds of particles were affected in the following way:

(i) All particles which showed no translatory movement at first began gradually (from about 20 gauss/cm.) to move towards the denser lines of force (i.e. towards the magnet).

(ii) Those particles which were at first attracted by the magnet continued to move in this direction at an increasing velocity.

(iii) Those particles which were at first repelled by the magnet gradually decreased their velocity, came to a standstill, and finally also moved towards the magnet with gradually increasing velocity.

The same results can be achieved by steadily approaching a small permanent bar magnet to the cell instead of using a coil with iron core and increasing the current as described above.

§ 4. TRACKS OF PARTICLES

Figure 3 is a general view showing tracks of some particles (the plane of the loop is vertical in the middle of the photograph). It is remarkable that the tracks often appear as dotted lines, which are in some cases enlarged to helix-like paths; this is illustrated by Figure 4, which shows the track of a single particle, and by Figure 5. Examination by means of a stereo-microscope proved these motions to be essentially rotatory, and not a sort of fluttering due to translatory motion.

Figure 6 shows for example the movements of a particle while the direction of the current in the loop was twice reversed. (Another particle showing only common photophoresis crossed the tracks of the first-mentioned particle.)

§ 5. FURTHER OBSERVATIONS

Ionization of the air in the cell by means of γ -rays did not influence the effect in any visible manner. Common radiometer forces play apparently only a small part in these movements, because the effect of common photophoresis was found to be only about 5 to 10% of the whole.

With care the same particle could be observed over a long period (half an hour and more), in course of which it did not lose the described property. Thus the explanation of the effect by assuming a jet-like emission of occluded gases caused by the impinging radiation, or as the result of a chemical action, cannot be accepted. Moreover, the velocity of the particles (in the direction of the magnetic field) is not constant, but increases with increase of field strength (up to about 120 gauss).

On an average the sensitivity of the particles in question is such that the method can be used for determining direction and intensity of weak magnetic fields (e.g. the geomagnetic one) when an accuracy of only ± 0.03 gauss is required.

§ 6. CONCLUSION

The described behaviour of the particles reminds one of equivalent experimental results, when electrically charged particles in inhomogeneous electric fields are observed, where the movements are described by the well-known equation

$$P = eE + f(\epsilon, K) \cdot \text{grad}(E^2), \quad \dots\dots (1)$$

in which P is the force acting on the particle, e the electric charge, E the electric field strength, and f a function of the dielectric constant ϵ and the shape K of the particle. (Conductivity is neglected.)

Analogy suggests a trial of the equation

$$P = qH + f(\mu, K) \cdot \text{grad}(H^2), \quad \dots\dots (2)$$

where H is the magnetic field strength and μ the magnetic permeability, but the coefficient of proportion q turns out to be a function of H , of intensity and composition of light. In addition to this a remanent dipole momentum has to be taken into account in (2), when ferromagnetic substances are used.

It is difficult to calculate values for q , because the shape of the particles (which can scarcely be assumed spherical as electron-optical photographs show) modifies Stokes' Law in a rather uncalculable manner. The values of q are approximately 10^{-9} (c.g.s.) and smaller, under the above-described circumstances. The effect is not restricted to irregularly shaped particles; it was found to be of the same order of magnitude with iron powder (precipitated from carbonyl iron), the particles of which consist of spheres.

ACKNOWLEDGMENT

I wish to thank Prof. Ehrenhaft for many discussions and stimulating suggestions. I am indebted to Mr. T. G. Hodgkinson for his advice and interest in this work.

REFERENCES

- BEISCHER, F., 1939, *Z. Elektrochem.*, **45**, 310.
 EHRENHAF, F., 1930, *C.R. Acad. Sci., Paris*, **190**, 263 ; 1942, *J. Franklin Inst.*, **233**, 235;
 1947, *J. Phys. Radium*, **8**, 55; 1948, *C.R. Acad. Sci., Paris*, **225**, 926.
 MAHL, H., 1940, *Z. techn. Phys.*, **21**, 17.
 SCHEDLING, J. A., 1949, *Phys. Rev.*, **76**, 843.

Dead Times of Self-Quenching Counters

By B. COLLINGE

George Holt Physics Laboratories, The University, Liverpool

Communicated by J. Rotblat ; MS. received 30th May 1949

ABSTRACT. Experiments on Geiger counters are described in which localization of the ion sheath was brought about by reducing the wire potential immediately after each count. The effect on the dead time is discussed. The dead time of a counter 30 cm. long was found to be 20 μ sec. and independent of the counting rate up to 1.8×10^6 counts per minute. The reduction of the dead time of short counters was also investigated.

§1. INTRODUCTION

It is an inherent property of the Geiger counter that each discharge is followed by an insensitive period or 'dead time', which may be of several hundred microseconds duration. This has been defined as "the time interval, after recording a count, . . . (during which) the counter is completely insensitive and does not detect other ionizing events occurring inside it" (Korff 1946). Counts occurring during this period are missed and the registered counting rate must be corrected for their loss. The form of this correction is uncertain; Blackman and Michiels (1948) quote three different formulae for the correction factor and show theoretically that only one of them is valid under normal conditions. Further difficulties are introduced by the dependence of the dead time on the operating conditions, which may include the counting rate.

One way of evading these difficulties is to work with such low counting rates that uncertainties in the correction to be applied may be neglected. Hence for 10% losses the counting rate must be limited to 10^4 counts per minute, if the paralysis time of the recording equipment is 600 μ sec. At higher counting rates the difficulty of applying an accurate correction increases rapidly. An empirical method of obtaining the true counting rate, when the dead time varies with it, has been described by Kohman (1945). This method is, however, rather laborious.

Another expedient is the addition to the recording equipment of an electronically determined paralysis time, much longer than the dead time of the counter. The purpose of the paralysis time is to reduce spurious counts and eliminate variations in the counter dead time. However, this does not solve the problem because the effective dead time is still not constant. This is so because a subsequent discharge of the counter, during the paralysis time, may prolong it by as much as the dead time.

For many experiments the upper limit set to the counting rate by these considerations is too low. It would be desirable, therefore, to obtain an improvement, either by reducing the dead time, or by ensuring its constancy. Thus, reducing the dead time would allow higher counting rates to be used for a given percentage of losses, while with a constant dead time larger corrections for losses could be safely applied. It would, of course, be most desirable to improve both factors.

The dead time may be made constant by the use of a circuit which lowers the wire potential below the threshold of the Geiger region for a few hundred microseconds immediately after each count. The counter is then truly 'dead' during

this constant period. There are several advantages of such a method of operation and Putman (1948) has shown that self-quenching counters operated in this way show improved plateaux and become free from temperature drift.

A number of attempts have been made to obtain a reduction of the dead time. The discharge occurring in a counter operating in the Geiger region is stopped or quenched by positive ions produced during the discharge (Montgomery and Montgomery 1940). These ions form a sheath round the wire, reducing the field and preventing further multiple ionization. A second discharge is not possible until the sheath has moved to a critical distance from the wire, such that the field is sufficiently restored. The time required for this movement of the ions has been identified by Stever (1942) as the dead time.

Sherwin (see Smith 1948) has pointed out that, since the ion sheath is formed within a few tenths of a millimetre of the wire, it should be possible to collect these ions on the wire, immediately after a discharge, by reversing the applied field. Under these conditions, the times required for the positive ions to reach the wire would be relatively short and a considerable reduction in the dead time should be obtained. Following this suggestion Simpson (1944) and Hodson (1948) have used circuits which switched the wire potential a hundred or so volts negative with respect to the cylinder for a few microseconds after each count and then returned it to the working value. Both these workers used long counters up to 60 cm. Dead times of 20 to 30 μ sec. were obtained.

If the mechanism by which the ion sheath spreads along the counter wire is considered, then another explanation of the reduction in dead times obtained by Simpson and Hodson is possible. Complete spreading of the ion sheath along the full length of the wire is a characteristic of the discharge of a counter operating in the Geiger region. Stever (1942) has shown that the ion sheath rarely spreads past a bead mounted on the wire in self-quenching counters. In a similar way, reducing the wire potential below the threshold of the Geiger region during the spread of the discharge will cause localization of the ion sheath.

A consideration of the velocity of spread of the ion sheath (Wantuch 1947, Hill and Dunworth 1946) suggested that reversing circuits similar to those used by Simpson and Hodson, when used with long counters, should cause the wire potential to fall below the threshold of the Geiger region before the sheath has had time to travel far along the wire. The remainder of the counter would then be sensitive as soon as the wire potential is restored to its operating value. Thus the reduction of the dead time obtained with a switching circuit may be due to two causes: collection of positive ions or localization of the ion sheath. Some of the experiments described in § 3 of this paper show that localization of the ion sheath does occur in long counters and that the explanation given by Simpson (1944) and Hodson (1948) of the reduction of the dead time obtained by them in their experiments was not complete. This conclusion (Collinge 1948) is in agreement with the work of Smith (1948), Hodson (see Smith 1948) and Elliot (1949). In this paper further experiments are described in which was determined the effect on the dead time of momentarily changing the wire potential.

§ 2. THE CIRCUIT

The circuit which is shown in Figure 1 was intended for experimental purposes only. The cylinder of the counter is earthed and the wire is connected to the anode of the switching valve, V_{14} . The anode current of this valve is normally cut off by negative bias on the control grid. The cathode is returned to a preset

The valves V_2 and V_4 amplify the pulses from the counter and feed them to the flip-flop formed by valves V_6 and V_9 . The 'switching time' is determined by the grid resistance of V_9 and may be varied from 1 to 10 $\mu\text{sec.}$; its amplitude, and

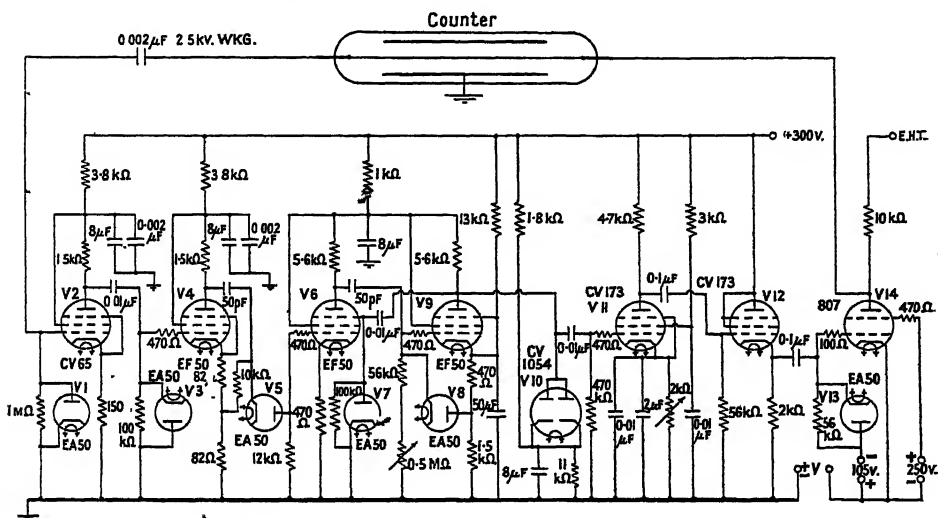


Figure 1. Geiger counter switching circuit.

The value of the switching voltage was measured by connecting the counter wire to the Y-plates of a cathode-ray oscillograph. The spot was normally biased off the screen by the high potential; each count caused it to be returned to the screen for the duration of the switching time and the switching voltage was read on a calibrated scale. This arrangement also enabled the drive to V_{14} to be adjusted so as to obtain square pulses.

(i) *Preliminary Work*

The operation of a counter with the circuit described is as follows. If localization of the ion sheath occurs, the region of the counter, not involved in a particular discharge, will be sensitive to a second ionizing particle as soon as the wire potential is restored and the circuit is ready to receive a second pulse. Thus using an oscillograph, triggered by pulses from a counter operated with the circuit, it should be possible to observe counts occurring at all intervals greater than the insensitive time of the circuit. This insensitive time is the sum of the switching

time, the recovery time of the circuit and the time required for the wire potential to return to a point on the plateau where the counter pulses are large enough to operate the circuit.

To test this, several counters of lengths varying from 5 to 30 cm. and containing argon plus alcohol mixtures to a total pressure of about 11 cm. were used. Each counter was connected in turn to the circuit and the anode potential of V_{14} adjusted to the operating voltage for the counter. Pulses from the cathode of V_{12} were used to trigger an oscilloscope, whilst the pulses occurring at the control grid at V_6 were fed to the Y-plates. A gamma ray source gave a convenient counting rate. The switching time was varied from 1 to 10 μ sec. and the switching voltage from plus to minus 500 volts.

Pulses were observed to occur on the cathode-ray oscillograph trace at all time intervals 2 or 3 μ sec. greater than the switching time. The recovery time of the circuit was about 3 μ sec.; this includes the time required for the wire potential to be restored. This experiment seems to confirm therefore that localization of the sheath was taking place.

(ii) Six-Cylinder Counter

Direct proof that localization of the ion sheath was occurring in these long counters was obtained using the six-cylinder counter shown in Figure 2. The cathode consists of six, separately insulated, identical cylinders, contained in a glass envelope. It was filled with 10 cm. of argon and 1 cm. of ethyl alcohol. The wire was connected to the anode of V_{14} and 1 M Ω resistances were connected

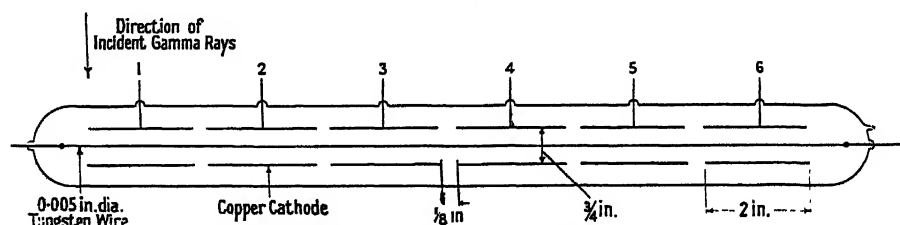


Figure 2. Six-cylinder counter.

Table 1

Counts per minute at cathode	No. 1	No. 2	No. 3
Circuit inoperative	1124	1150	1131
Circuit operative	918	97	43

between each of the cylinders and ground. A discharge, occurring in a particular cathode, caused a positive pulse to appear across the associated resistance. This counter was arranged so that a collimated beam of gamma rays was directed, at right angles to the wire, on to No. 1 cathode.

With a switching time of 1 μ sec. and zero switching voltage, the counting rates, at cathodes No. 1, 2 and 6, were determined by connecting an amplifier, discriminator and scale of 100 to each of the cathodes in turn. Readings were also obtained with the switching circuit made inoperative by disconnecting the condenser feeding the control grid of V_{14} . The results obtained are given in Table 1, which shows the average counting rate registered, at the three cathodes, after correcting for the background counting rate.

(iii) Two-Window Counter

The counter shown in Figure 3 was used to obtain further confirmation in a similar way. It consists of a brass cathode with a mica window at each end, parallel to the wire. An RCA photo-multiplier tube was mounted opposite each window so that the light, produced during the discharge in that region of the counter, fell on the light sensitive surface. The voltage pulses, occurring at the collecting electrode, were recorded on a scaler via an amplifier and discriminator.

As before, the switching voltage was adjusted to zero and the switching time to $1\ \mu\text{sec}$. Experiments were performed with either a collimated source of γ -rays

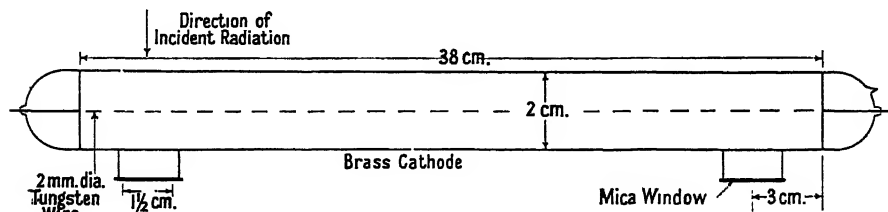


Figure 3. Two-window counter.

Table 2

Counts per minute	Window near source			Window remote from source		
	γ -rays	β -rays	β -rays	γ -rays	β -rays	β -rays
Circuit inoperative	2051	3614	1687	1693	3736	1674
Circuit operative	2087	3526	1568	220	42	16

Table 3

Counter	Six-cylinder		Two-window		
Type of radiation	γ	γ	γ	β	β
Distance between centres of detectors (cm.)	5	22.5	32	32	32
% of discharges remote from source	10.5	4.5	13.5	1	1

directed at one end of the cylinder or with a beam of β -rays from ^{32}P passing into the counter at right angles to the wire, through an aluminium window. This window, not shown in Figure 3, was in the brass wall opposite one of the mica windows.

In Table 2 the number of counts per minute registered by the two photo-tubes are shown for both β and γ -rays. The figures are average values and have been corrected for the background.

(iv) Discussion

The results of Tables 1 and 2 taken with the switching circuit operative are condensed in Table 3. The bottom line shows the percentage of discharges detected at various distances along the counters. These results show that localization of the ion sheath does occur.

A higher percentage of the discharges was localized when β -rays were used; this is explained by incomplete collimation and scattering of the γ -rays. Even so, the results for the six-cylinder counter show that about 90% of the discharges was confined to less than 5 cm. of the wire.

§ 4. DETERMINATION OF THE DEAD TIME OF A LONG COUNTER

(i) *The Paired Source Method*

It was stated in § 3 that the dead time of a long counter obtained from a triggered oscillograph was equal to the insensitive time of the switching circuit. The dead time determined by this method is independent of the length of the counter provided it is long enough for localization of the ion sheath to occur. Thus with a switching time of $1 \mu\text{sec.}$, dead times of 3 or $4 \mu\text{sec.}$ were found for several counters of lengths greater than 5 cm. Although the result is consistent with Korff's definition of dead time, it is of no value for the correction of counting losses. This is because the whole of the counter is not sensitive to further counts immediately after the dead time. Regions of the counter where the ion sheath is localized remain insensitive for the normal dead time imposed by the ion sheath. The average losses depend on a number of factors which include the length of the counter. The true dead time must be obtained by a method giving the losses directly.

A convenient way of determining the true dead time involves the use of two sources. The average counting rate is noted when the sources are presented to the counter, separately and together.

If the registered counting rate is n_r then the true counting rate n_t is given by

$$\frac{n_r}{n_t} = \frac{1}{1 + n_t \tau}, \quad \dots\dots(1)$$

where τ is the dead time.

From equation (1) Beers (1942) has derived a convenient formula for the dead time. Neglecting the background counting rate this may be written

$$\tau = \frac{\Delta}{2n_1 n_2} + \frac{n_{12} \Delta^2}{8n_1^2 n_2^2}, \quad \dots\dots(2)$$

where $\Delta = n_1 + n_2 - n_{12}$ and the counting rates n_1 , n_2 and n_{12} are derived from N_1 , N_2 and N_{12} the total counts registered by the separate and combined sources during the intervals t_1 , t_2 and t_{12} respectively.

The error in the dead time can be calculated approximately. Suppose that errors in time measurements are neglected and the errors in the total counts arranged to be of the same order by making $N_1 \simeq N_2 \simeq N_{12} \simeq N$ say, then $t_1 \simeq t_2 \simeq 2t_{12} \simeq t$ for roughly equal sources.

Then from equation (2) it can be shown that $\delta\tau$, the probable error in τ , is given by equation (3) where δN is the probable error in N .

$$\delta\tau = \sqrt{\left(\frac{3}{2}\right) \frac{t \delta N}{N^2}}. \quad \dots\dots(3)$$

The probable error P_τ due to statistical variations in the total number of counts is obtained by putting $\delta N = \frac{2}{3}\sqrt{N}$

i.e.

$$P_\tau = \sqrt{\left(\frac{2}{3}\right) \frac{t}{N\sqrt{N}}},$$

from which it follows that

$$t = \frac{16}{3} \frac{(t_{12})^3}{(N_{12})^3} \frac{1}{P_\tau^2}. \quad \dots\dots(4)$$

Then for chosen values of P_r equation (4) gives the necessary counting period t for various values of t_{12}/N_{12} . Some values of t in minutes are given in Table 4, row 2, for $P_r = 3 \mu\text{sec}$.

Table 4

Counting rate n_{12} (sec^{-1})	1,000	2,000	3,000	10,000	30,000
Minimum counting period t necessary to ensure $P_r < 3 \mu\text{sec}$. (min.)	10	<2	<2	$\ll 2$	$\ll 2$
P_r , the probable error in τ for $\alpha = \frac{1}{2}\%$ (μsec .)	12	6	5	1	0.3

Errors due to uncertainties in the wire potential may often be neglected when finding the dead time of a counter operated in the normal way. This is not true when a switching circuit is used which draws an average current which may be of the order of a milliampere, because of the difficulty of ensuring the same counter voltage at various counting rates.

If the probable error in N , due to setting the counter voltage supply is δN , then $\delta N/N$ is the fractional error, α say. Then from equation (3)

$$P_r = \sqrt{6\alpha t_{12}/N_{12}}. \quad \dots\dots (5)$$

Row 3 of Table 4 shows the values of P_r for $\alpha = \frac{1}{2}\%$ for various values of N_{12}/t_{12} .

(ii) Experimental Details and Results

The counter used consisted of a brass cathode, 32 cm. long and 2 cm. inside diameter, with a 0.2 mm. tungsten wire, and filled with 10 cm. of argon and 1 cm. of ethyl alcohol. In order to minimize the effect of positive ion collection the switching voltage was set to zero and the switching time at $1 \mu\text{sec}$. A scale of 10^4 , with a paralysis time of about $5 \mu\text{sec}$. was used to record the counts. Two gamma-ray sources were used at a distance adjusted to give the desired counting rate. Readings were taken with the sources in several positions and the dead time determined as a function of the counting rate. A similar series of measurements was made with the circuit inoperative. The counting interval was always chosen from Table 4 so that the error due to statistical variation of the total counts could be neglected in comparison with α .

The dead times obtained are shown in Figure 4 where they have been plotted against the counting rate. The values of the counting rate, obtained with the two sources together, have been plotted along the x -axis after correcting for losses. Curve (a) was taken with the circuit not working and (b) with it working.

(iii) Discussion

As the experiment was arranged so that the statistical errors could be neglected we need only consider errors due to setting of the H.T. The experimental arrangement enabled its value to be determined to about 5 volts. The slope of the counter characteristic at the wire potential used was 0.1% per volt, determined under the conditions for the dead time measurements. Thus in this experiment $\alpha = \frac{1}{2}\%$. Hence the values of P_r applicable to the dead times obtained are not greater than those given in row 3 of Table 4. These errors apply to both graphs. The slope of 0.1% per volt is typical of the characteristic of a counter of this type, taken with a short paralysis time.

If the statistical dead time depends strongly on the counting rate, the value obtained from equation (2) loses its significance. Its meaning, in fact, is defined by this equation. This point has been discussed by Curran and Rae (1947). The values given in curve (a) Figure 4 are therefore somewhat artificial. The results show however the large changes which occur when the counting rate is high and emphasize the difficulties of correcting for large losses.

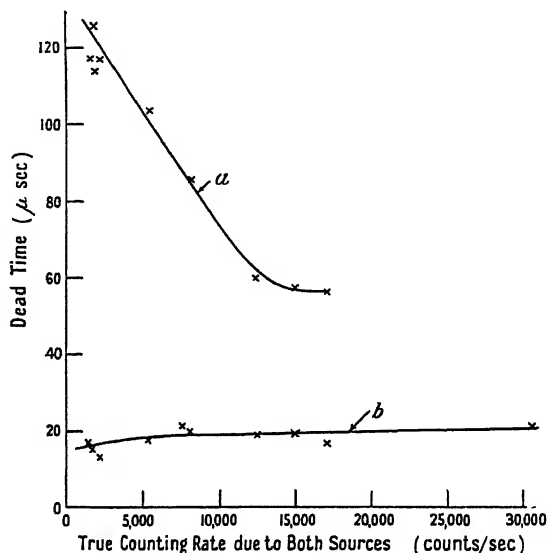


Figure 4. Dead time as a function of counting rate for a long counter.
(a) Circuit inoperative; (b) Circuit operative.

The results shown in curve (b) obtained with the circuit operative show that not only is the dead time considerably reduced but its value is constant and independent of the counting rate within the experimental error. This is true up to counting rates of nearly 2×10^6 counts per minute. It would be reasonable to correct for losses greater than 10% under these conditions. For a statistical dead time of 20 μ sec. the losses would be 10% at a counting rate of 3×10^5 counts per minute.

§ 5. EXPERIMENTS WITH SHORT COUNTERS

In the experiment described in this section G.E.C. type G.M.2 counters were used; they are single ended argon-alcohol β counters. The tungsten wire is about 2 cm. long. Such a counter was connected to the circuit and a gamma-ray source used to provide a suitable counting rate.

The dead time of such a counter was observed on a triggered oscillograph in the manner previously described. Dead times considerably greater than the switching time and dependent on the circuit parameters were obtained, as would be expected if the wire is too short for localization of the ion sheath to occur. To show that localization was not occurring, V_s was used as a simple discriminator. The cathode was biased with a positive potential which was increased until triggering occurred only with pulses of maximum amplitude. This is a simple way of introducing a time delay to ensure that triggering occurs only when the ion sheath has had time to spread along the whole length of the wire. No change in the dead time was observed, showing that the ion sheath always spreads along the whole of the wire.

As complete spreading occurs in these counters it was considered reasonable to observe the dead time on a triggered oscillograph. The values obtained are of qualitative interest and are related to the values which would be obtained using the paired source method.

The dead times obtained for one of the counters which had a dead time of 105 $\mu\text{sec.}$ when the switching circuit was inoperative, have been plotted as a function of the switching time on Figure 5. Several curves are shown for various

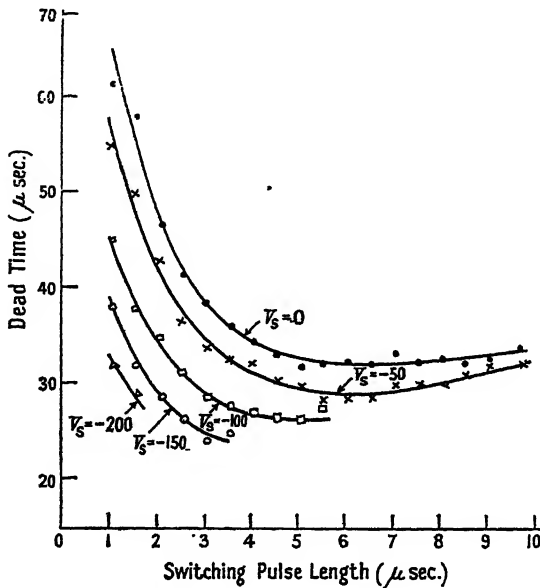


Figure 5. Variation of the dead time of a short counter with length of switching pulse. V_s is the switching potential.

values of the switching time. For these results the anode load of V_{14} was changed from 10 k Ω to 50 k Ω .

The curves show an expected initial decrease in dead time as the switching time increases. This is consistent with a progressive increase in the number of positive ions collected on the wire. As the minimum dead time is not the same for all the curves and is very much greater than the switching time, it is concluded that part only of the ion sheath is collected. This is in agreement with the observed distribution of pulse amplitudes which shows an increase in amplitude from zero at the dead time to a maximum value. It is presumed that the decreased dead time corresponds to the time taken for the reduced ion sheath to move to a critical distance from the wire at which Geiger avalanches can occur.

As the switching time is increased the rate of collection of positive ions decreases. This is because it is more difficult to collect ions remote from the wire away from the very high field region. The rate of collection decreases until the reduction in dead times obtained just equals the increase in the switching time. Thus the curves should have a minimum and a positive slope, approaching unity, for large values of the switching times. This conclusion is in agreement with the curves shown.

Some of the curves of Figure 5 do not include dead times for the full range of switching times. This is because for these conditions the cathode-ray oscillograph showed that very many spurious counts were occurring.

The reason why a reduction in dead times is obtained even when the switching voltage is negative has been discussed by Smith (1948). The charge induced by the sheath on the wire leads to large fields near the wire sufficient to cause collection of positive ions.

§ 6. COUNTER CHARACTERISTICS

In general no significant changes in counter characteristics have been observed when using counters with the switching circuit. One counter had a somewhat longer plateau when the switching voltage was minus 150 volts. Three commercial gamma-counters had plateaux with slopes which became progressively worse as the switching voltage was made more negative. All the counters tested produced larger numbers of spurious counts when the switching voltage was very negative.

In order to reduce the chance of producing spurious counts, it is probably an advantage not to collect positive ions at the wire but to reduce the dead time of long counters by localizing the ion sheath. In this case a simpler circuit could be used to produce a very short negative pulse at the wire with an amplitude of 300 or 400 volts. It would be desirable that such a circuit should trigger on very small pulses in order to ensure that the time available for the ion sheath to spread is as small as possible. Not only is it desirable to limit the extent of the ion sheath, in order to obtain a short dead time, but also because the life of the counter is increased by reducing the number of positive ions involved in each discharge. Elliot (1949) describes some experiments in which the charge collected at the wire is shown to be reduced when a switching circuit is used and finds that under these conditions the life of the counter is prolonged.

ACKNOWLEDGMENTS

The author wishes to express his gratitude to the Department of Scientific and Industrial Research for the grant which made this work possible and to thank Sir James Chadwick for his interest and Dr. J. Rotblat and Dr. J. D. Craggs for many helpful discussions.

REFERENCES

- BEERS, Y., 1942, *Rev. Sci. Instrum.*, **13**, 72.
 BLACKMAN, M., and MICHIELS, J. L., 1948, *Proc. Phys. Soc.*, **60**, 549.
 COLLINGE, B., 1948, *Nature, Lond.*, **162**, 853.
 CURRAN, S. C., and RAE, E. R., 1947, *Rev. Sci. Instrum.*, **8**, 871.
 ELLIOT, H., 1949, *Proc. Phys. Soc. A*, **62**, 369.
 HILL, J. M., and DUNWORTH, J. V., 1946, *Nature, Lond.*, **158**, 833.
 HODSON, A. L., 1948, *J. Sci. Instrum.*, **25**, 11.
 KOHMAN, T. P., 1945, *Atomic Energy Declassified Report*, MDCC. 905.
 KORFF, S. A., 1946, *Electron and Nuclear Counters* (New York: Van Nostrand).
 MONTGOMERY, C. G., and MONTGOMERY, D. D., 1940, *Phys. Rev.*, **57**, 1030.
 PUTMAN, J. L., 1948, *Proc. Phys. Soc.*, **61**, 312.
 SIMPSON, J. A., 1944, *Phys. Rev.*, **66**, 39.
 SMITH, P. B., 1948, *Rev. Sci. Instrum.*, **19**, 453.
 STEVER, H. G., 1942, *Phys. Rev.*, **61**, 38.
 WANTUCH, E., 1947, *Phys. Rev.*, **71**, 464.

Breakdown in Cold-Cathode Tubes at Low Pressure

By F. G. HEYMANN

University of the Witwatersrand, South Africa

Communicated by G. I. Finch ; MS. received 17th May 1949

ABSTRACT. When, in cold-cathode discharge tubes, electrode spacings and surface areas deviate from the ideal of infinite parallel planes, it is found that the breakdown voltage between electrodes increases. This is explained as a decrease in the effective η of the gas due to loss of electrons and positive ions by diffusion to the walls of the container. The loss factor per unit potential difference is shown theoretically to be inversely proportional to the field strength and tube radius, although this is not fully verified by experiment.

Paschen curves obtained experimentally for potassium and nickel cathodes in argon and in a neon-argon mixture at low pressures are shown and, from these, values of γ for the two cathode surfaces are obtained as a function of E/p_0 . The apparatus used for measuring breakdown voltage is described.

Earlier theories of statistical and formative time delays are extended to cover the case of a rising overvoltage and also the case where the primary electrons appear in bursts as with ionization by α -particles. The shape of the statistical distribution curve is an indication of whether the primary electrons have been produced singly or in bursts. The overvoltage ΔV of breakdown due to formative lag and the rate of rise of the uniformly increasing applied voltage V are found to bear the relation $\Delta V \propto \sqrt{(dV/dt)}$. This has been experimentally verified.

§ 1. INTRODUCTION

THIS subject has been treated by quite a number of authors and experimental work has been carried out with various electrode materials and gases. The mechanism of breakdown is fairly well understood but the factor γ which represents the secondary mechanisms has not yet been fully explored because of the difficulty of direct determination. In the following pages some possibilities are set down in this connection.

When electrodes of finite dimensions are used in breakdown measurements, the special case of the similarity principle known as Paschen's law does not hold as will be shown later and an approximate theoretical approach will be attempted.

§ 2. INFINITE PARALLEL PLANE ELECTRODES

The simplest breakdown gap from a theoretical point of view is that between infinite parallel planes giving uniform field strength.

When all primary electrons are emitted from the cathode, the current density increases in the field direction according to the following relation (Penning and Druyvesteyn 1940, Rogowski 1940):

$$j_e = j_{ke} \exp \{ \eta (V - V_0) \}, \quad \dots\dots(1)$$

where V is the potential with respect to the cathode.

If all primary electrons are due to ionization in the gas by an external agent, the current relation is

$$j_e = j_s [\exp \{ \eta (V_a - V_0) \}] / \eta V. \quad \dots\dots(2)$$

In the steady state the positive ion current density at the cathode which corresponds to the electron current density of equation (1) is

$$j_{ki} = j_{ke} [\exp \{ \eta (V_a - V_0) \} - 1]. \quad \dots\dots(3)$$

The probability that a positive ion which reaches the cathode causes emission of an electron may be represented by γ_{ia} and the probability that the electron will not return to the cathode by diffusion by γ_{ib} . The probability that a positive ion will give rise to an active electron is thus: $\gamma_i = \gamma_{ia}\gamma_{ib}$.

Let there be θ photons created in the gas for each ionization and let a fraction f_r fall on the cathode. The average probability that a photon will give rise to an active electron may be represented by $\gamma_{ra}\gamma_{rb}$ in analogy to the case of positive ions. The secondary emission factor due to photons is then (Penning and Druyvesteyn 1940)

$$F_r\gamma_r = (\theta f_r)(\gamma_{ra}\gamma_{rb}).$$

The factor f_r depends on the geometry of the gap and on photon absorption in the gas. In the present case, if absorption is negligible, $f_r = \frac{1}{2}$.

θ is a function of the fundamental ratio E/p_0 which is the main factor in discharge at low pressure (Penning and Druyvesteyn 1940, Penning 1933).

For metastable atoms an expression similar to that for photons will hold. (Penning and Druyvesteyn 1940): $F_m\gamma_m = (\psi f_m)(\gamma_{ma}\gamma_{mb})$. ψ is analogous to θ , both factors increasing with decreasing E/p_0 . f_m depends on diffusion of the metastable atoms and is therefore related to the geometry of the gap, the pressure of the gas, and is also a function of time until the steady state is reached when metastable distribution between infinite parallel planes is (Rogowski 1940, Newton 1948)

$$m(x) = j_{ke} \frac{\psi}{\eta ED} [\{\exp\{\eta(V_a - V_0)\} - 1\} \frac{x}{d} - \{\exp\{\eta(V_x - V_0)\} - 1\}] \quad \dots\dots (4)$$

and

$$f_m = \frac{1}{\eta(V_a - V_0)} - \frac{1}{\exp\{\eta(V_a - V_0)\} - 1} \quad \dots\dots (5)$$

The total secondary emission factor which relates the number of secondary electrons entering the discharge per second to the number of ionizations occurring in the gas per second is then (Penning and Druyvesteyn 1940)

$$\gamma = \gamma_i + F_r\gamma_r + F_m\gamma_m \quad \dots\dots (6)$$

The factors γ_{ia} , γ_{ra} and γ_{ma} are similar since they represent the efficiency of secondary emission. It has been shown by Penning (1930) and Oliphant (1929, 1930) that the kinetic energies of ions and metastable atoms have little effect on these factors when the energies are low and of the order encountered in low pressure discharge. The excitation energies are therefore responsible for secondary emission (Meili 1945) which means that the γ_a factors should be substantially constant over a large range of field strength and pressure.

The factors γ_{ib} , γ_{rb} and γ_{mb} represent the effect of back-diffusion which has been treated by several writers (Hertz 1927, Pose 1928-9, Thomson 1928, Loeb 1939, Meili 1945).

The general expression derived is (Pose 1928-9)

$$\gamma_b = \left\{ 1 + \frac{3\bar{V}}{4\lambda_0 E/p_0} \log \left(\frac{V_0}{\bar{V}} + 1 \right) \right\}^{-1} \quad \dots\dots (7)$$

This is based on the assumption that only elastic collisions occur in the distance λ_0 between the cathode and the anode which is approximately true.) It is evident from the

It has thus been indicated that all the factors in equation (6) depend directly on E/p_0 or are constants, except f_m , in the case of infinite parallel planes. If the contribution of the metastable atoms or the variation in f_m is not large, γ will approximately be a function of E/p_0 alone.

In the steady state f_m is a constant and if this state is approximately reached before final breakdown takes place γ will be a function of E/p_0 .

The breakdown relation may be written in the form (Penning and Druyvesteyn 1940)

$$V_B = V_0 + \frac{1}{\eta} \log \left(1 + \frac{1}{\gamma} \right). \quad \dots\dots (8)$$

The factors V_0 and η have been determined experimentally (Penning and Druyvesteyn 1940, Kruithof 1940) and are functions of E/p_0 for pure gases.

Thus the breakdown voltage V_B will be a function of E/p_0 in the case of infinite parallel plane electrodes.

Since the field is uniform

$$p_0 d = \frac{V_B}{E/p_0}, \quad \dots\dots (9)$$

so that V_B will be a function of $p_0 d$ alone, which is a statement of Paschen's law.

The minimum breakdown voltage corresponds roughly with the maximum value of η (Penning and Druyvesteyn 1940) which for the rare gases is approximately $1/3 V_i$.

The factor γ varies enormously amongst various metals, being the greatest for the alkali metals and smallest for iron according to measurements made up to the present in the rare gases.

§ 3. FINITE PLANE ELECTRODES

When plane electrodes are used in measurements they have to be very large compared with their distance of separation to approximate to infinite planes. Loss of ionization products occur at the boundaries of the gap due to diffusion, and the main cause of increase of breakdown voltage is the decrease in the effective η as a result of the loss of electrons and ions. The change in γ will not be very great since only the factors F_r and F_m are reduced. When these factors become important at low values of E/p_0 this may not be true but changes in γ do not affect the breakdown voltage so much, since it occurs in a logarithmic function.

The greater part of the decrease in η is due to electron loss because of the large mean energy and consequent high rate of diffusion.

In the breakdown or Townsend region at low pressure, space charge is unimportant (Penning and Druyvesteyn 1940) and it may be assumed that the diffusion of electrons and ions depends on their respective concentrations.

At a distance x from the cathode the potential in a uniform field is $V_x = Ex$ and the potential difference over a distance dx is $dV = E dx$.

The increment of electron current density due to ionization by collision is $\eta j_e dV$ whilst the loss by diffusion may be represented by $\sigma_e j_e dV$.

Thus:

$$dj_e = (\eta - \sigma_e) j_e dV, \\ j_e = j_{ke} \exp \{ (\eta - \sigma_e)(V_x - V_0) \}. \quad \dots\dots (10)$$

The increment of ions is $\eta j_e dV$ and the loss $\sigma_i j_i dV$. Therefore

$$di = -\eta j_e dV + \sigma_i j_i dV.$$

Substituting for j_e and solving

$$j_i = A \exp \{ \sigma_i (V_x - V_0) \} - \frac{\eta j_{ke}}{\eta - \sigma_e - \sigma_i} \exp \{ (\eta - \sigma_e)(V_x - V_0) \}.$$

Setting $\sigma = \sigma_e + \sigma_i$ and at the anode $V_x = V_a$, $j_i = 0$,

$$j_i = j_{ke} \frac{\eta \exp \{ \sigma_i (V_x - V_0) \}}{\eta - \sigma} [\exp \{ (\eta - \sigma)(V_a - V_0) \} - \exp \{ (\eta - \sigma)(V_x - V_0) \}].$$

.....(11)

The positive ion current at the cathode is the same as at $V_x = V_0$,

$$j_{ki} = j_{ke} \frac{\eta}{\eta - \sigma} [\exp \{ (\eta - \sigma)(V_a - V_0) \} - 1].$$

.....(12)

Breakdown will occur when $\gamma j_{ki} = j_{ke}$,

$$\frac{\gamma \eta}{\eta - \sigma} [\exp \{ (\eta - \sigma)(V_B - V_0) \} - 1] = 1,$$

$$(\eta - \sigma)(V_B - V_0) = \log \left(1 + \frac{\eta - \sigma}{\gamma \eta} \right) \simeq \log \left(1 + \frac{1}{\gamma} \right) - \frac{\sigma}{\eta(1 + \gamma)},$$

$$\simeq \eta(V_{B0} - V_0) - \frac{\sigma}{\eta(1 + \gamma)},$$

$$V_B - V_{B0} = \Delta V; \quad V_{B0} - V_0 - \frac{1}{\eta(1 + \gamma)} = V_R,$$

$$\Delta V = \frac{\sigma}{\eta - \sigma} V_R; \quad \frac{\sigma}{\eta} = \frac{\Delta V}{\Delta V + V_R}.$$

.....(13)

These are approximate expressions linking up the increase in breakdown voltage with the loss of electrons and ions due to diffusion.

The factor σ_e is difficult to determine theoretically and the following derivation gives mainly a qualitative result:

If the discharge is confined to the diameter of the electrodes of radius R by a glass envelope of slightly larger diameter and the electron concentration near the glass wall is n_w , then the number striking unit area on the glass wall per second is $q_w = n_w \bar{c}_e / 4$. The average electron concentration in the tube is j_e / eu_e (j_e is average current density).

Let n_w be related to this density by a factor a such that $n_w = a j_e / eu_e$. The loss per elemental distance dx is then

$$2\pi R q_w dx = 2\pi R \frac{a j_e \bar{c}_e}{4 eu_e} dx,$$

now

$$u_e = K_e \frac{E}{p_0} = \text{const.} \times \frac{\lambda p_0}{\sqrt{\bar{V}_e}} \frac{E}{p_0}$$

and

$$\bar{c}_e \propto \sqrt{\bar{V}_e}.$$

Therefore the loss factor per unit potential difference is

$$\sigma_e = \text{const.} \times \frac{a \bar{V}_e p_0}{\lambda p_0} \frac{I_e}{E ER}.$$

.....(14)

This is probably the main part of σ and if σ_1 is assumed to lead to a similar relation, the factor σ/η is approximately given by

$$\frac{\sigma}{\eta} = \text{const.} \times \frac{\bar{V}_e a p_0 I}{\lambda p_0 \eta E E R}. \quad \dots (15)$$

At constant E/p_0 the only variables are E and R and with constant R , σ/η should be inversely proportional to E if the factor a does not vary.

Mierdel and Steenbeck (1937) approach the problem differently by assuming the average life of an electron or ion to be $T = R^2/(2.4)^2 D$. On this assumption σ/η becomes inversely proportional to $E^2 R^2$.

In the experimental section the relative merits of the two derivations will be discussed with reference to the actual results.

§ 4. MEASUREMENT OF BREAKDOWN VOLTAGE

A diagram of the apparatus together with a sketch of the type of experimental tube used are shown in Figure 1. In measuring breakdown voltage, a steady

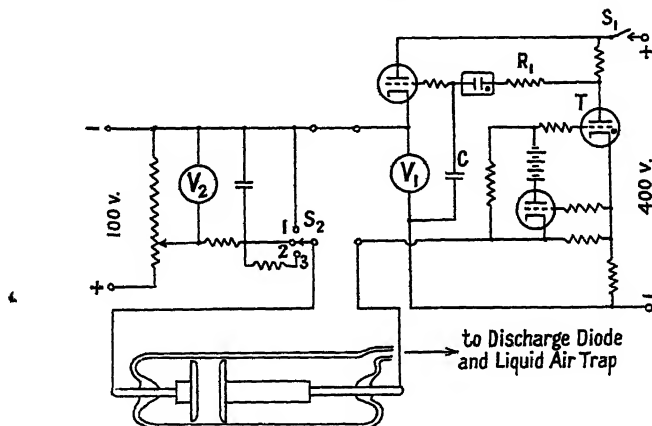


Figure 1. Apparatus for measuring breakdown potential and a typical experimental tube.

potential about 100 to 200 volts less than the expected breakdown voltage was applied through the potentiometer shown and the magnitude was indicated on voltmeter V_2 . On closing switch S_1 , the capacitance C charged through resistor R_1 and the potential across V_1 rose at the same rate, thus adding a further potential across the tube. When breakdown occurred, thyatron T fired, thus stopping the charging of C and applying an inverse voltage in series with V_1 and V_2 to reduce or extinguish the current through the tube, depending on the voltage. V_1 could read up to 240 volts and the inverse voltage was about 170 volts.

The final value attained by V_1 together with V_2 is the maximum voltage across the tube just before a discharge passes. We may thus define breakdown potential as measured by this equipment as the maximum voltage between the electrodes prior to the negative slope transition region.

The cathodes of the experimental diodes were illuminated sufficiently in order to reduce statistical spread and as will be seen later (Figure 8), the overvoltage at the rate of rise of V_1 of 40 volts per second is negligible.

The experimental tube consisted of two mild steel electrodes enclosed by a close-fitting hard glass tube. One of the electrodes was made to slide on a $\frac{1}{8}$ in.

diameter tungsten rod and could be moved by using an external magnet. The tube was evacuated and baked at 400° c. for some time and the electrodes were brought to a dull red heat by eddy-current heating. In almost all the cases potassium-activation was used and a good, reproducible potassium layer could be deposited on the electrodes by vacuum distillation.

Rare gases were used in the spectrally pure state and further purified by keeping in a vessel in which a continuous discharge was maintained between potassium-activated electrodes. Condensable vapours were removed in a double liquid-oxygen trap and as a last safeguard a potassium-activated diode passing a continuous discharge was permanently connected in the inlet to the main tube.

Three sizes of tube (dimensions in centimetres) were used:

Tube	Electrode diameter	Tube diameter	Maximum electrode separation
1	1.5	1.6	2.5
2	2.8	3.0	2.6
3	6.0	6.5	3.75

Breakdown voltage was measured as a function of electrode separation at various pressures in argon and in a mixture of neon and argon which contained 6.6% argon (potassium-activated cathode).

Some representative curves for tube 2 are shown in Figure 2. Breakdown voltages are plotted as functions of p_0d and it is evident that the curves at the lower pressures do not follow Paschen's law.

In Figure 3 the curves of breakdown voltage in argon are plotted for all three tubes as functions of E/p_0 , assuming uniform field distribution. As was to be expected, the differences between curves are greatest for the smallest tube and least for the largest, for the same differences in pressure. At higher pressures the curves tend to merge in all cases and the curves for the various tubes show good correspondence at the pressures at which the losses by diffusion become small.

In order to determine the effect of the nearness of the glass wall to the edges of the electrodes, a tube with 1.5 cm. diameter electrodes and a 3 cm. diameter glass envelope was constructed. The values of breakdown voltage were not greatly different from those in tube 1 at the lower pressures, and it was therefore decided that the glass wall had only a small effect.

In order to check the validity of the theoretical expression for σ/η given in equation (15), the curves in Figure 4(a) were drawn, using equation (13) to determine σ/η . The theory predicts that σ/η should be inversely proportional to E at constant E/p_0 and it has been found that there is reasonable correspondence as shown by the experimental points in Figure 4(a).

These points were calculated at various values of E/p_0 (from 50 to 200 volt/cm.mm.), and it seems that for each tube a single factor of proportionality will represent all the points satisfactorily. The correspondence is not exact and partly due to experimental and graphical errors. Furthermore the theory is only approximate.

The correspondence amongst the three tubes is not so good, the theory predicting that the slopes of the lines in Figure 4(a) should be inversely proportional to the electrode diameters, which is not the case. It is probable that there is a further factor which has been overlooked in the theoretical derivation or it may be that the factor α in equation (15) is a function of R .

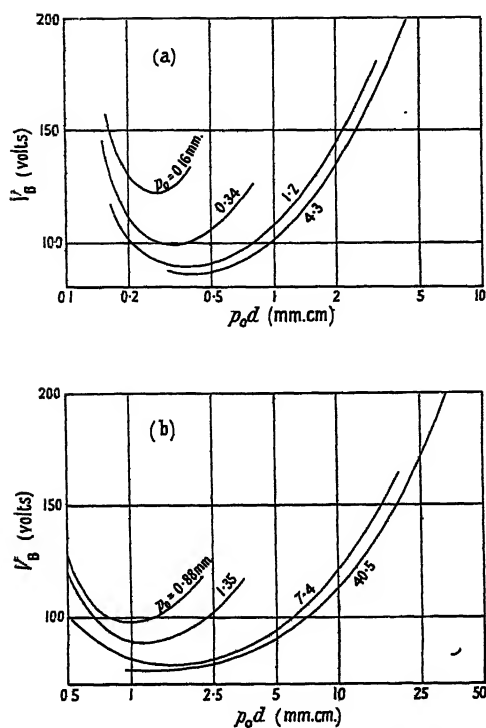


Figure 2. Breakdown potential curves with electrodes of 2.8 cm. diameter at various pressures :
(a) Pure argon; (b) 93.4% neon, 6.6% argon.

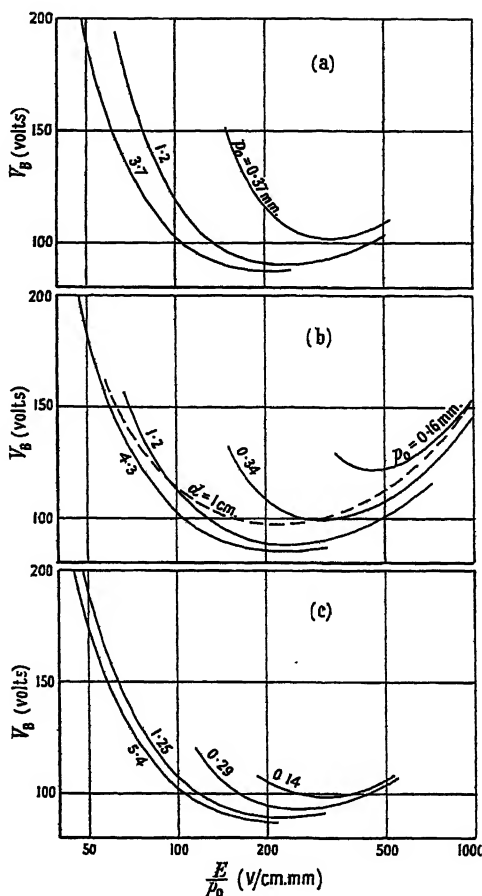


Figure 3. Breakdown potential curves as functions of E/p_0 in pure argon. Electrode diameter :
(a) 1.5 cm.; (b) 2.8 cm.; (c) 6.0 cm.

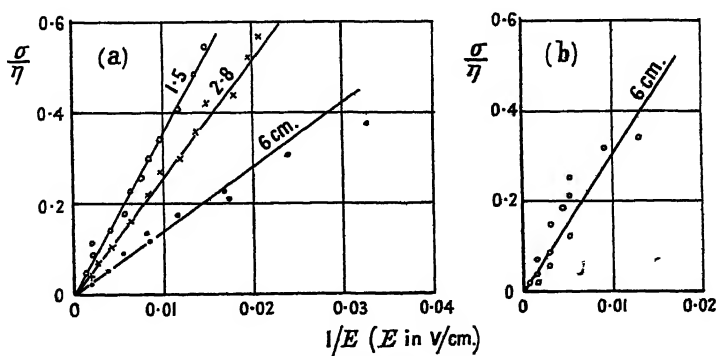


Figure 4. Experimental points representing $\sigma/\eta = \Delta V / (\Delta V + V_R)$.
(a) Argon : potassium-activated electrodes (electrode diameters in cm.).
(b) Argon : nickel electrodes.

There seems to be a further objection to the theoretical expression in that the factor σ/η for nickel electrodes of 6 cm. diameter derived from Figure 4(b) does not correspond to that for potassium-activated electrodes.

The theoretical approach according to Mierdel and Steenbeck (1937) applies even less than the theory derived in this paper and cannot therefore explain the increase in breakdown voltage.

Figures quoted by McCallum and Klatzow (1934) for increase in breakdown voltage between nickel electrodes in argon are much higher than those found in this series of experiments and must be due to an additional cause not present here.

The increase of breakdown potential with small electrodes and at low pressures makes it necessary to use large electrode diameter, small spacing and high pressure to obtain a true Paschen curve. The values of breakdown potential measured on the tube of 6 cm. diameter will correspond fairly well with the values between infinite parallel planes except perhaps at high values of p_0d in argon where the pressure was about 50 mm. Hg and the separation up to 3.75 cm.

The measured curves for nickel and potassium over an extended range of p_0d are shown in Figure 5. The curve for potassium is easily reproducible and on

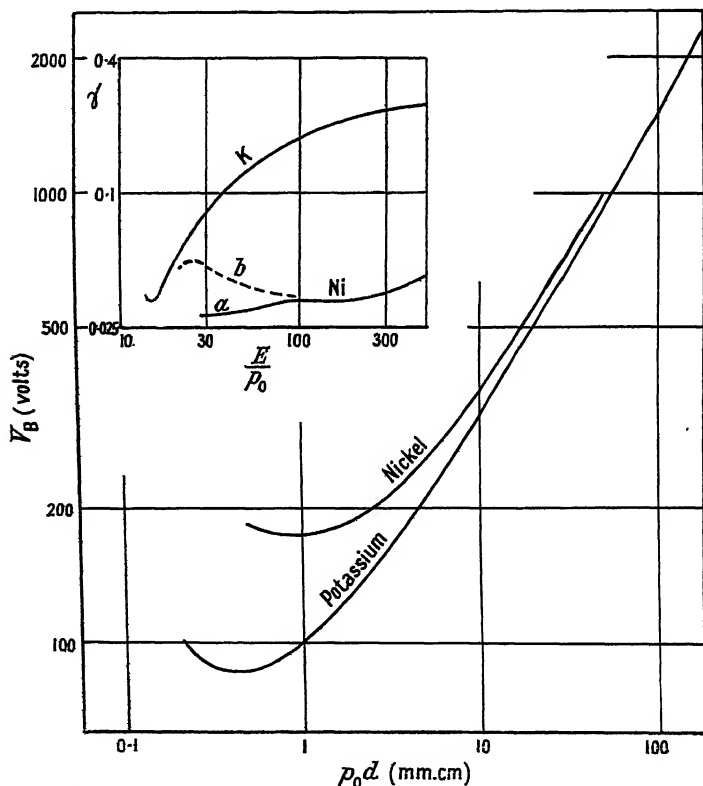


Figure 5. Experimental breakdown potential curves in argon with potassium and nickel cathodes. Inset: γ -curves as functions of E/p_0 , for potassium and two different nickel cathodes.

using each electrode as cathode in turn, very close correspondence between the breakdown potentials occurred.

The nickel was in the form of an electro-deposit on the steel electrodes and although the treatment of both electrodes was exactly similar, the breakdown potentials with each as cathode in turn differed appreciably.

The factor γ was calculated from the breakdown relation in equation (8), although high accuracy is not attained by this method.

The values of V_0 and η were taken from Penning and Druyvesteyn (1940) and Kruithof (1940). Figure 5 (inset) shows γ -curves for potassium and nickel as functions of E/p_0 . Curve *a* for nickel corresponds with the p_0d breakdown voltage curve shown. Curve *b* is for the other electrode and merges with *a* at higher values of E/p_0 .

The curve of γ for potassium shows a form similar to that to be expected if the γ_b factors were predominant, the other factors being approximately constant. At very low values of E/p_0 the curve tends to rise again as has been found for other materials by others (Penning and Druyvesteyn 1940) and is probably due to the increasing importance of θ and ψ as E/p_0 decreases.

The curves for nickel show that the metastable atom and photon action become important below a value of E/p_0 of 100 volt/cm.mm. and to a different extent for the two electrodes, depending on their sensitivity to this action. Again this confirms previous results (Penning and Druyvesteyn 1940), but the full explanation is still obscure.

This illustrates the fact of the difficulty of the interpretation of γ and the importance of the purity of the cathode material.

§ 5. TIME DELAY EFFECTS IN BREAKDOWN

There are two main causes of time delays in breakdown which will be treated separately and then combined. The delays involved when constant overvoltage is applied have already been dealt with by others (von Laue 1925, Zuber 1925, Schade 1936-7, Hertz 1937) and will serve as a basis for relations to be derived for the case of linearly increasing overvoltage.

The statistical time lag is due to insufficient primary electrons and formative lag comes about as a result of the time taken for positive ions to move to the cathode from their points of origin.

§ 6. STATISTICAL TIME DELAY

Primary electrons, which are necessary to initiate breakdown, appear in a random fashion due to some external agent and therefore introduce time delays, even when the applied potential is adequate for breakdown. Let formative time delay be neglected for the time being. Von Laue (1925) has shown that the probability that breakdown will take place after a time t_s measured from the instant of applying an adequate potential, is given by

$$f(t) = \exp \left(-q_e \int_0^{t_s} W dt \right). \quad \dots\dots (16)$$

In the case of plane parallel electrodes, Hertz (1937) has shown that W , the probability that one electron will initiate discharge, is related to the factor M as follows:

$$M = \gamma [\exp \{ \eta (V_s - V_0) \} - 1] = \frac{\gamma \log (1 - W)}{\log (1 - \gamma W)}.$$

At the normal value of breakdown potential $M=1$ and $W=0$.

With constant overvoltage, i.e. when M is constant and greater than unity, W is constant and the probability of breakdown becomes

$$f_1(t) = \exp (-q_e W t_s) = \exp (-t_s/t_0). \quad \dots\dots (17)$$

This relation has been verified experimentally by Zuber (1925).

In order to extend the theory to the case of rising overvoltage, it is necessary to make some approximations:

If γ and W are small at small overvoltage ΔV , then

$$M \simeq -\frac{1}{W} \log(1 - W) \simeq 1 + \frac{1}{2}W.$$

Using a method of Schade (1938), it follows approximately that

$$\begin{aligned} M &= \gamma [\exp \{(\alpha_B + \Delta\alpha)(d - d_0)\} - 1] \\ &\simeq \gamma [\exp \{\alpha_B(d - d_0)\} - 1] + \gamma \exp \{\alpha_B(d - d_0)\} \frac{\Delta\alpha}{\Delta E} \Delta E(d - d_0) \\ &\simeq 1 + \frac{\partial\alpha}{\partial E} \Delta V. \end{aligned} \quad \dots\dots(18)$$

Thus

$$W \simeq 2 \frac{\partial\alpha}{\partial E} \Delta V \simeq 2 \frac{\partial\alpha}{\partial E} \frac{dV}{dt} t = \frac{\partial W}{\partial t} t. \quad \dots\dots(19)$$

Therefore the probability of breakdown after a time t_s with the applied potential rising linearly from the normal breakdown value, is

$$f_2(t) = \exp \left\{ -\frac{1}{2} q_0 (\partial W / \partial t) t_s^2 \right\} = \exp \left\{ -\frac{1}{2} (t_s / t_m)^2 \right\}. \quad \dots\dots(20)$$

The probability of breakdown occurring in an interval between t_s and $t_s + dt_s$ is given by

$$f_2'(t) dt_s = -\frac{t_s}{t_m} \exp \left\{ -\frac{1}{2} (t_s / t_m)^2 \right\} d(t_s / t_m) \quad \dots\dots(21)$$

with a maximum occurring at $t_s = t_m$.

The functions f_1 and f_2 are represented in Figure 6 in terms of their characteristic time constants, t_0 and t_m respectively.

Function f_2 is only valid for small values of overvoltage and cannot be applied where the overvoltage becomes large. It is further restricted if the electrons do not occur singly but in b bursts of r electrons each. The factor q_0 must then be replaced by br and W by $W_r = 1 - (1 - W)^r$. Whilst W (and therefore ΔV) is very small this does not differ much from f_2 since $b(\partial W_r / \partial t) = br(\partial W / \partial t) = q_0(\partial W / \partial t)$, but W_r soon approaches unity and the function goes over into a form similar to f_1 . A better approximation at larger overvoltages may be made by assuming W to have the form $W = 1 - \exp(-t/T)$, where T is approximately determined by finding the value of ΔV_T which corresponds to $W = 1 - e^{-1}$, giving $T = \Delta V_T / (dV/dt)$.

The probability of breakdown for electrons occurring singly then becomes

$$f_{3a}(t) = \exp [-q_0 \{t_s - T(1 - \exp(-t_s/T))\}]. \quad \dots\dots(22)$$

If the average number of electrons formed simultaneously is r with b bursts per second,

$$f_{3b}(t) = \exp [-b \{t_s - (T/r)(1 - \exp(-rt_s/T))\}]. \quad \dots\dots(23)$$

The probabilities of breakdown in any particular interval may be found by differentiation and it is found that maxima occur at the following values corresponding to equations (22) and (23) respectively:

$$f_{3a}: t_s/T = \log \frac{2q_0 T}{2q_0 T + 1 - \sqrt{4q_0 T + 1}}, \quad \dots\dots(24)$$

$$f_{3b}: t_s/T = \frac{1}{r} \log \frac{2bT/r}{2bT/r + 1 - \sqrt{4bT/r + 1}}. \quad \dots\dots(25)$$

Some general curves representing f_{3a} and f_{3b} are shown in Figure 7(a). In these $q_e T = brt = 2$ is constant and r is varied. For any other value of $q_e T$ the probability is just raised to a proportional power.

It is evident that already with bursts of 10 electrons at a time the curve f_{3b} approaches closely the exponential curve $\exp(-bt_a)$.

Thus the transition is fairly rapid from one shape of probability curve to the other. The form of the statistical distribution curve will indicate the way in which the primary electrons are produced. Where these electrons are due to emission from the cathode, they are independent of each other and therefore the distribution represented by f_{3a} is to be expected.

When the primary electrons are produced by α -particles, they will occur in groups, each particle giving rise to a whole train of electrons which are simultaneous for most practical purposes, so that f_{3b} should be applicable.

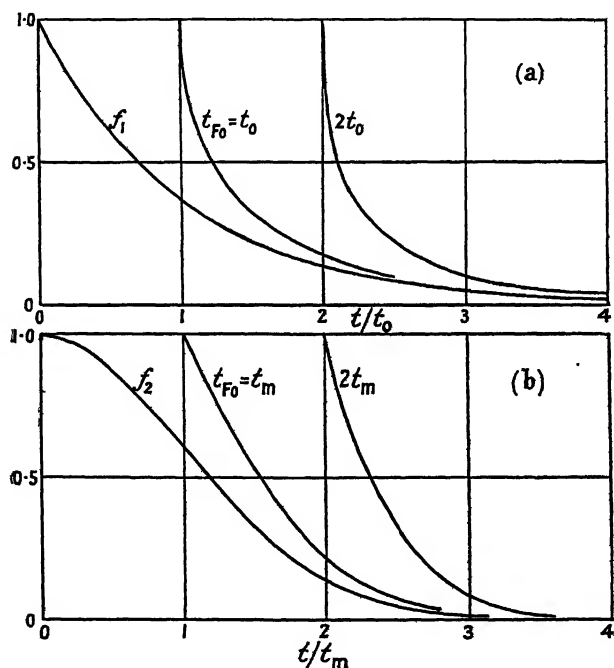


Figure 6. Statistical time delay probability curves f_1 and f_2 and curves of combined statistical and formative time delays.

Ionization by other types of radiation does not usually give rise to a closely spaced group of electrons and the distribution will tend to approximate to the case of single electrons.

Electrons which are produced in the body of the gas are less effective than electrons emitted from the cathode. The probability function W is based on the latter and in the case of gas ionization it may be shown that W must be divided by $\eta(V - V_0)$ to compensate for this fact.

Some interesting experiments have been carried out in order to determine statistical delay time when a discharge tube was kept in total darkness away from radioactive sources. The only external source of ionizing radiation could have been cosmic rays and the very feeble natural radioactivity of materials.

The experimental tubes used were small sealed-off diodes with potassium activated cathodes and argon filling at 18 mm. Hg pressure. These were of the parallel plane electrode type, electrodes spaced 0.06 in. apart and $\frac{1}{4}$ in. in diameter.

It was found that primary electrons far in excess of those to be expected from the above-mentioned causes were present in the tube when the breakdown potential was measured at intervals of three seconds. On increasing the interval between measurements to 90 seconds, the primary electrons were much less, showing that the short discharge occurring at each measurement had some influence on the subsequent time delay. Between measurements the inter-electrode field was reduced to zero.

Representative curves are shown in Figure 7(b). The distribution curve for measurements at 3 second intervals yields a calculated rate of primary electron

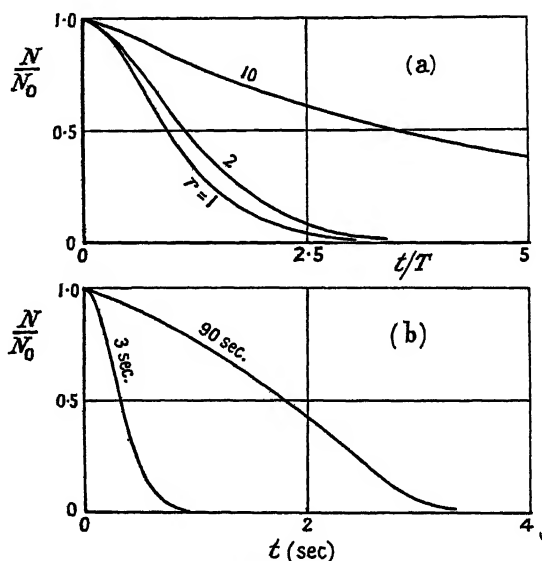


Figure 7.

- (a) Statistical time delay probability curves f_{ga} and f_{gb} .
 (b) Experimental statistical distribution curves with different intervals between measurements. (Rate of voltage rise—40 v/sec.)

production of about 5 per second. With 90 second intervals the rate is about one electron in two seconds, whereas the number which will be produced naturally is of the order of one per minute.

This effect may be attributed to continued emission from the cathode after extinction of a discharge, as Paetow (1939) has shown. He found that this continued emission was still appreciable after 50 seconds and that it was probably due to small specks of impurity on the cathode which acquired charges during discharge and afterwards caused electron emission due to the high fields set up. It is probable that even on high purity potassium such specks may exist and give rise to this persistent emission which was still noticeable when the interval between small measurement discharges was 90 seconds.

Paetow found that the emission was very large immediately after stopping a discharge, partly due to metastable action which became negligible after about a second and thereafter due to this after-effect only.

§ 7. FORMATIVE TIME DELAY

When statistical time delay is reduced to negligible proportions by sufficient primary electrons, the discharge will not take place instantaneously on application of sufficient overvoltage, since it takes a definite time for the current to increase to the point at which conduction is appreciable.

This formative time delay has been dealt with by Schade (1936-37) for the case of constant overvoltage, yielding the following expression:

$$t_F = t_1 \frac{M}{M-1} \log \left\{ \frac{1 + (M-1)j_{ke}/j_{k0}}{M} \right\}. \quad \dots\dots(26)$$

The main cause of the delay in current increase is the average transit time of positive ions from their points of creation to the cathode. The primary electron current density at the cathode is j_{k0} and the final cathode electron current density j_{ke} and their ratio which is included in the logarithmic term has a minor effect on the formative time delay.

The expression for formative delay has not previously been derived for the case of rising applied potential and will be developed below.

By the method used before in equation (18) it is possible to show that

$$M-1 \simeq \frac{\partial \alpha}{\partial E} \Delta V_F = \frac{\partial \alpha}{\partial E} \frac{dV}{dt} t_F.$$

For small overvoltages ($M-1$) will be small and the time delay will be approximately inversely proportional to the average value of ($M-1$), according to equation (26).

If the overvoltage rises linearly, the average overvoltage will be half of the final value reached

$$(M-1)_{av} \simeq \frac{1}{2} \frac{\partial \alpha}{\partial E} \frac{dV}{dt} t_{F0},$$

and from equation (26)

$$t_{F0} = \frac{2t_1}{(\partial \alpha / \partial E)(dV/dt)t_{F0}} \log \{1 + (M-1)_{av} j_{ke}/j_{k0}\},$$

giving

$$t_{F0} = \left[\frac{1}{dV/dt} \frac{2t_1}{\partial \alpha / \partial E} \log \{1 + (M-1)_{av} j_{ke}/j_{k0}\} \right]^{\frac{1}{2}}. \quad \dots\dots(27)$$

The final overvoltage will therefore be

$$\Delta V_m = \left[\frac{dV}{dt} \frac{2t_1}{\partial \alpha / \partial E} \log \{1 + (M-1)_{av} j_{ke}/j_{k0}\} \right]^{\frac{1}{2}}. \quad \dots\dots(28)$$

The terms in the logarithmic function can cause very little variation so that it follows approximately that

$$t_{F0} \propto (dV/dt)^{-\frac{1}{2}}, \quad \Delta V_m \propto (dV/dt)^{\frac{1}{2}}. \quad \dots\dots(29)$$

On deriving the expression more rigorously from Schade's (1936-37) differential equation, the result is

$$t_{F0} = \left[\frac{1}{dV/dt} \frac{2t_1}{\partial \alpha / \partial E} \log \left\{ j_{ke}/j_{k0} \left(1 + \left(\frac{2}{\pi} \frac{\partial \alpha}{\partial E} \frac{dV}{dt} t_i \right)^{\frac{1}{2}} \right)^{-1} \right\} \right]^{\frac{1}{2}}. \quad \dots\dots(30)$$

There is not much difference when compared with equation (27) which may therefore be used.

The limiting formative time delay when $(M-1)$ tends to zero becomes (Schade 1936-37)

$$\lim_{M \rightarrow 1} t_{F0} = t_i j_{k0} / j_{k0} \dots \dots (31)$$

The average distance through which positive ions move to the cathode may be found by integrating over the whole discharge gap, yielding

$$d_{av} = d \left\{ \frac{e^{\alpha d}}{e^{\alpha d} - 1} - \frac{1}{\alpha d} \right\} \dots \dots (32)$$

This is for plane parallel electrodes so that the average transit time is

$$t_1 = \frac{p_0 d}{K_1 V_a} d_{av} = \frac{d_{av}}{K_1 E / p_0} \dots \dots (33)$$

In the case of a concentric diode in which the anode radius R_a is much smaller than the cathode radius R_c , the average transit time is approximately

$$t_1 = \frac{R_c - R_a}{2 K_1 E_c / p_0} \dots \dots (34)$$

where E_c is the field strength at the cathode.

There is some experimental confirmation of the equation (28) in that the linear portions of measured curves have slopes which correspond well with the calculated values.

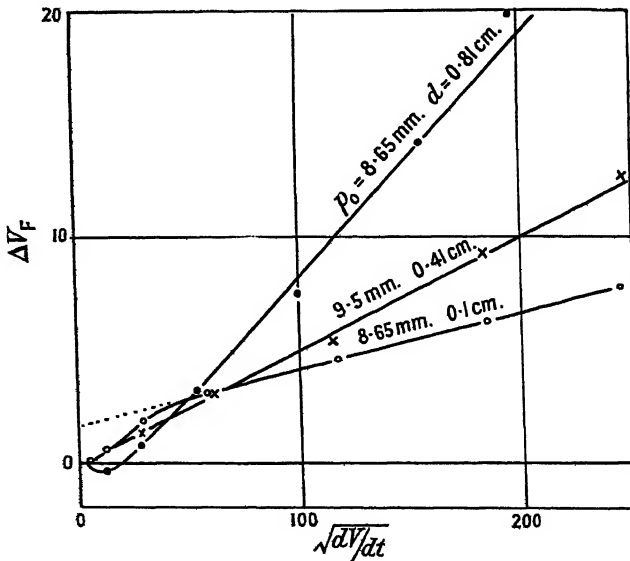


Figure 8. Experimental formative overvoltage curves at various values of p_0 and d . The linear portions correspond well with the theoretical calculations. dV/dt expressed in v/sec.

Figure 8 shows typical curves from which it is evident that in general the initial portions of the curves do not conform to the theory but that the curves have a constant slope over most of the range of rate of voltage rise.

The expected slope of an experimental curve is $\Delta V_m / \sqrt{(dV/dt)}$, from equation (28).

Assuming $\log \{1 + (M-1)_{av} j_{ke}/j_{k0}\}$ to be approximately constant at 18 and taking $\partial\alpha/\partial E$ from Figure 9 (Kruithof 1940) the following table may be drawn up for argon and 6 cm. diameter potassium-activated electrodes to compare experiment with theory:

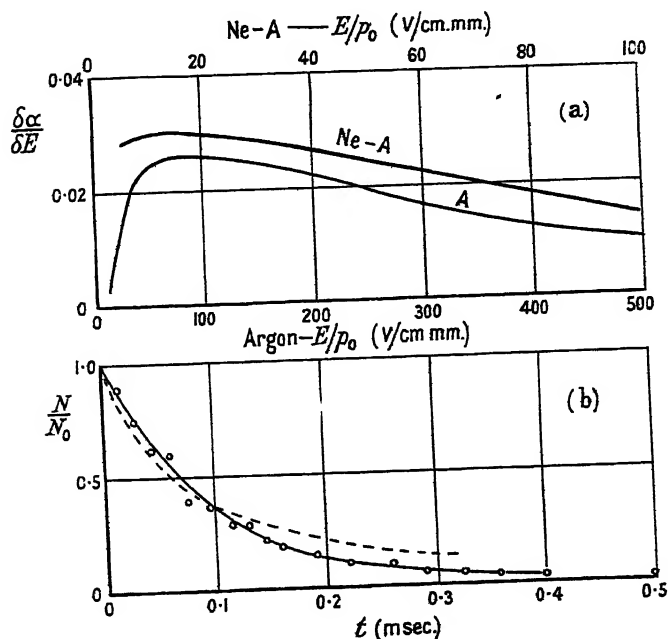


Figure 9. (a) The factor $\partial\alpha/\partial E$ for argon and for a mixture of 93.4% neon with 6.6% argon. (b) Experimental distribution curve showing the effect of combined statistical and formative delays.

p_0	Experimental			Theoretical Slope
	d	t_1	Slope	
4.56	0.104	0.17	0.023	0.017
	0.195	0.53	0.031	0.028
	0.405	1.74	0.050	0.049
	0.8	4.95	0.086	0.083
9.5	0.1	0.29	0.025	0.020
	0.2	0.89	0.035	0.035
	0.41	2.74	0.065	0.064
8.65	0.1	0.27	0.025	0.020
	0.2	0.84	0.036	0.034
	0.41	2.58	0.050	0.061
	0.81	7.05	0.114	0.107

Mobility of argon positive ions at 1 mm. Hg = 2.1×10^8 cm/sec/volt/cm.

The theory thus finds some justification in practice except at small overvoltage where more factors seem to be present than those which have been taken into account. The order of overvoltage calculated is correct and the theory may therefore be applied with reasonable accuracy.

This formative delay is not a variable quantity in a fully symmetrical discharge gap and may therefore be allowed for in cases where it becomes important.

§ 8. COMBINED STATISTICAL AND FORMATIVE DELAYS

If average statistical and formative delays are of the same order, an expression may be derived in order to take this into account for rising potential.

With statistical time delay, a certain time elapses before the discharge is initiated and therefore the overvoltage will reach a value ΔV_s before the formative period begins.

Let the increase of overvoltage during the formative period be ΔV_F . Then the average overvoltage during this period is $\Delta V_s + \frac{1}{2}\Delta V_F$.

The formative time delay is therefore given by equation (26) as

$$t_F = \frac{1}{\Delta V_s + \frac{1}{2}\Delta V_F} \cdot \frac{2t_i}{\partial \alpha / \partial E} \log \{1 + (M-1)_{av} j_{ke} / j_{k0}\},$$

$$\Delta V_F = \frac{dV}{dt} t_F = \frac{\Delta V_{F0}^2}{\Delta V_s + \frac{1}{2}\Delta V_F},$$

where V_{F0} is the formative overvoltage when $V_s = 0$.

The solution to the above equation is

$$\Delta V_F = -\Delta V_s + \{\Delta V_s^2 + \Delta V_{F0}^2\}^{\frac{1}{2}}. \quad \dots\dots (35)$$

The total overvoltage is

$$\Delta V = \Delta V_s + \Delta V_F = \{\Delta V_s^2 + \Delta V_{F0}^2\}^{\frac{1}{2}}, \quad \dots\dots (36)$$

$$t = \{t_s^2 + t_{F0}^2\}^{\frac{1}{2}}. \quad \dots\dots (37)$$

The statistical probability curves f_1 and f_2 are modified as shown in Figure 6 when the formative time delays are of the same order as the statistical delay time constants t_0 and t_m (see equations (17) and (20)).

A typical statistical distribution of points for combined delay times is shown in Figure 9(b), as measured on a small concentric diode containing 15 mm. argon and about $1 \mu\text{gm}$. γ -equivalent radioactive material to provide primary ionization. The measured formative time delay at a rate of voltage rise of 3×10^4 volts/sec. was about 0.25 millisecond. The solid curve was fitted at the ordinate e^{-1} to correspond to a modified f_2 probability curve with $t_{F0} = 2t_m$. The broken curve was fitted at the ordinate e^{-1} and corresponds to a modified f_1 probability curve with $t_{F0} = t_0$. No better fit is possible with this latter type of distribution and it thus seems as if the f_2 type of distribution given in equation (20) gives a very good fit. This is however not what would be expected if the α -particles created groups of electrons containing about 50 electrons each as they should do in this case.

The number of primary electrons occurring per second on calculating with the f_2 type of distribution is about 4×10^4 per second, which is of the order of the number of α -particles to be expected per second.

Theoretically the curve should follow the f_1 type of distribution which in this case does not seem to be confirmed. Other tubes show similar curves which means that there may be a further factor to take into account.

ACKNOWLEDGMENTS

Acknowledgments are due to Messrs. Ferranti Limited, who have provided the facilities for carrying out this work, and to the Admiralty who have sponsored the development and have given permission for the publication of this paper.

REFERENCES

- HERTZ, G., 1927, *Z. Phys.*, **46**, 177 ; 1937, *Ibid.*, **106**, 102.
 HUXFORD, W. S., and ENGSTROM, R. W., 1940, *Phys. Rev.*, **58**, 67.
 KRUTHOF, A. A., 1940, *Physica*, **7**, 519.
 VON LAUE, M., 1925, *Ann. Phys.*, *Lpz.*, **76**, 261.
 LOEB, L. B., 1939, *Fundamental Processes of Electrical Discharge in Gases* (New York : John Wiley).
 MCCALLUM, S. P., and KLATZOW, L., 1934, *Phil. Mag.*, **17**, 291.
 MEILI, E., 1945, *Helv. Phys. Acta*, **18**, 79.
 MIERDEL, G., and STEENBECK, M., 1937, *Z. Phys.*, **106**, 311.
 NEWTON, R. R., 1948, *Phys. Rev.*, **73**, 570.
 OLIPHANT, M. L. E., 1929, *Proc. Roy. Soc. A*, **124**, 228 ; 1930, *Ibid.*, **127**, 373.
 PAETOW, H., 1939, *Z. Phys.*, **111**, 770.
 PENNING, F. M., 1930, *Proc. K. Ned. Akad. Wet. Amst.*, **33**, 841 ; 1933, *Physica*, **5**, 268.
 PENNING, F. M., and DRUYVESTEYN, M. J., 1940, *Rev. Mod. Phys.*, **12**, 87.
 POSE, H., 1928-9, *Z. Phys.*, **52**, 428.
 ROGOWSKI, W., 1940, *Z. Phys.*, **115**, 257.
 SCHADE, R., 1936-7, *Z. Phys.*, **104**, 487 ; 1938, *Ibid.*, **108**, 353.
 THOMSON, J. J., 1928, *Conduction of Electricity through Gases* (Cambridge : University Press), **1**, 466.
 ZUBER, K., 1925, *Ann. Phys.*, *Lpz.*, **76**, 231.

The Behaviour of Multiple Circuit Magnetrons in the Neighbourhood of the Critical Anode Voltage

BY W. E. WILLSHAW AND R. G. ROBERTSHAW

Research Laboratories of The General Electric Company, Limited, Wembley, Middx.

MS. received 30th June 1949

ABSTRACT. It is pointed out that the mechanism of operation of the multiple circuit magnetron oscillator in the region of minimum magnetic field and voltage, where the efficiency is commonly assumed to approach zero, should approach that of an oscillator of the travelling wave tube type, providing that a cathode of suitable size is used. Useful efficiencies should thus be obtainable under these conditions.

Details of experiments are given in which an electronic efficiency of 12% was obtained at a wavelength of 3 cm. at values of magnetic field and voltage several times lower than those used for high efficiency operation. The mode of operation was determined by the value of the magnetic field, a given mode being maintained over a range of magnetic field of the order of 8%. The anode voltage was about 70% of the critical value.

The experimental results generally support the hypothesis and suggest that the minimum voltage regime should be of extreme importance for work at the highest radio frequencies.

§ 1. INTRODUCTION

THE main characteristics of performance of the high efficiency multiple circuit magnetron oscillator are well known. Operation in one mode starts at minimum values of magnetic field strength and anode voltage, and continues with increasing efficiency as these parameters are increased. In the higher efficiency region it is assumed that electrons interacting favourably with the oscillating field leave the main electron cloud, which extends only part of the way from cathode to anode, and being gradually focused into tight bunches finally arrive at the anode with a velocity which is the same as that of a rotating wave component of the anode field. On this assumption a very good approximation to the observed efficiency is derived directly, i.e. $\eta = 1 - \frac{1}{2} m \Omega^2 a^2 / eV$.

where Ω is the rotating wave angular velocity, a the anode radius, and V_a the anode voltage (Willshaw and Rushforth 1946). Change of mode, giving a change in Ω may occur at any value of magnetic field strength as V_a is changed in order to change current. For the lower efficiency region, where the anode voltage and magnetic field strength are much lower, the main electron cloud extends much nearer to the anode, so that electrons leaving it do not have time to be formed into bunches before reaching the anode, and to acquire the velocity of the rotating wave component. Thus the observed efficiency is much lower than that given by the above expression.

§2. BEHAVIOUR NEAR THE CRITICAL VOLTAGE

At the limiting lower values of magnetic field and anode voltage the electron cloud extends quite close to the anode surface, and there is experimental evidence to show that in the absence of oscillations electrons may describe circular paths which are coaxial with the anode cylinder (Wassermann 1948). Now the electron velocity at radius r is $eH/2m(1 - b^2/r^2)$, H being the magnetic field strength, and b the cathode radius. If the radius of the cathode is not too large relative to that of the anode, this velocity will vary very little in the neighbourhood of the anode. The variation in electron density in the neighbourhood of the anode is similarly small, being proportional to $1 + (b/r)^4$. Conditions are thus similar to those in a travelling wave tube (Kompfner 1947), in that there is a stream of electrons of given velocity and density passing across the mouths of several cavities and therefore capable of interacting with the fields due to oscillations in these cavities. By direct analogy there is then the possibility of oscillations being generated when the electron velocity close to the anode is equal to the velocity of the wave generated in the anode circuit. This condition will obtain when $(eH/2m)(1 - b^2/r_m^2) = \Omega = (2eV_a/m)^{1/2}$ where r_m is the maximum radius reached by the space charge. A change in mode of oscillation corresponding to a change in Ω should result when H or V_a are correspondingly changed, this being in distinct contrast to the behaviour in the high efficiency region. We shall call this the 'minimum voltage' regime of operation.

§3. EXPERIMENTAL PERFORMANCE IN THE MINIMUM VOLTAGE REGIME

Some experiments have recently been carried out in the minimum voltage regime with a magnetron having 28 cavities of the 'rising sun' form (Hollenberg, Kroll and Millman 1948) with a circuit ratio of 2.25/1, anode diameter of 16 mm., cathode diameter of 3 mm., and wavelength of π mode of 3.31 cm. Figure 1. This valve, fitted with a much larger cathode (11.5 mm. dia.) had previously been designed and used for high power operation with an anode voltage of the order of 15 kv., magnetic field of 2,500 oersteds and efficiency of 20%. Figure 2 shows the relationship between anode voltage and mean current for the minimum voltage regime for a range of magnetic field values, and contours of constant power output are included. The valve was operated with a 50 c/s. voltage applied to the anode through a 5,000 ohm series resistance so that current passed for approximately one fifth of the time. The corresponding wavelength of operation is shown together with the mode number n of the wave component with which the electron is interacting. The mode number is physically the number of

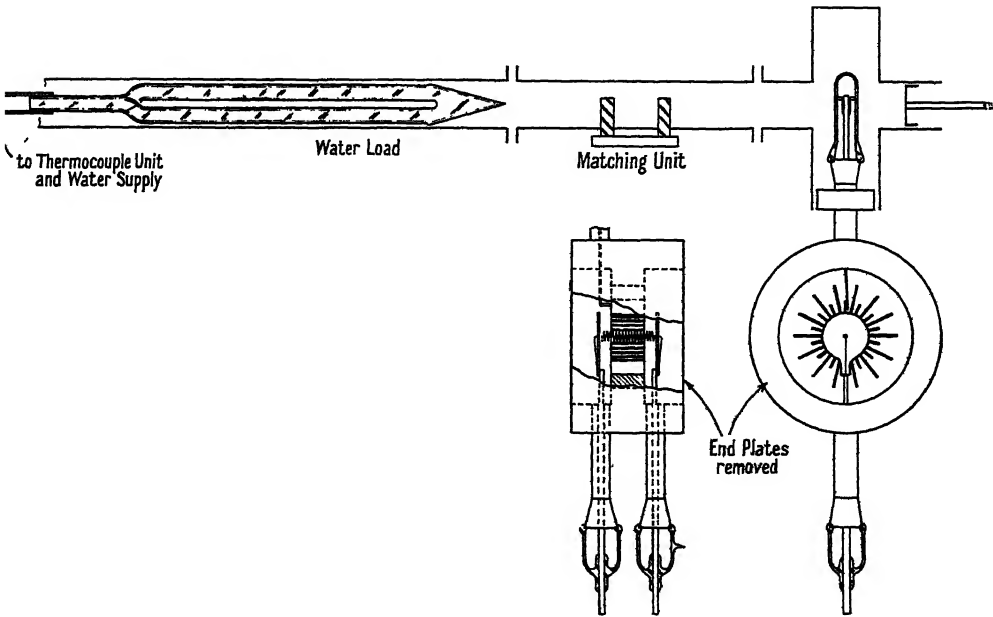


Figure 1. Rising sun magnetron for 3.3 cm. wavelength, and water load.

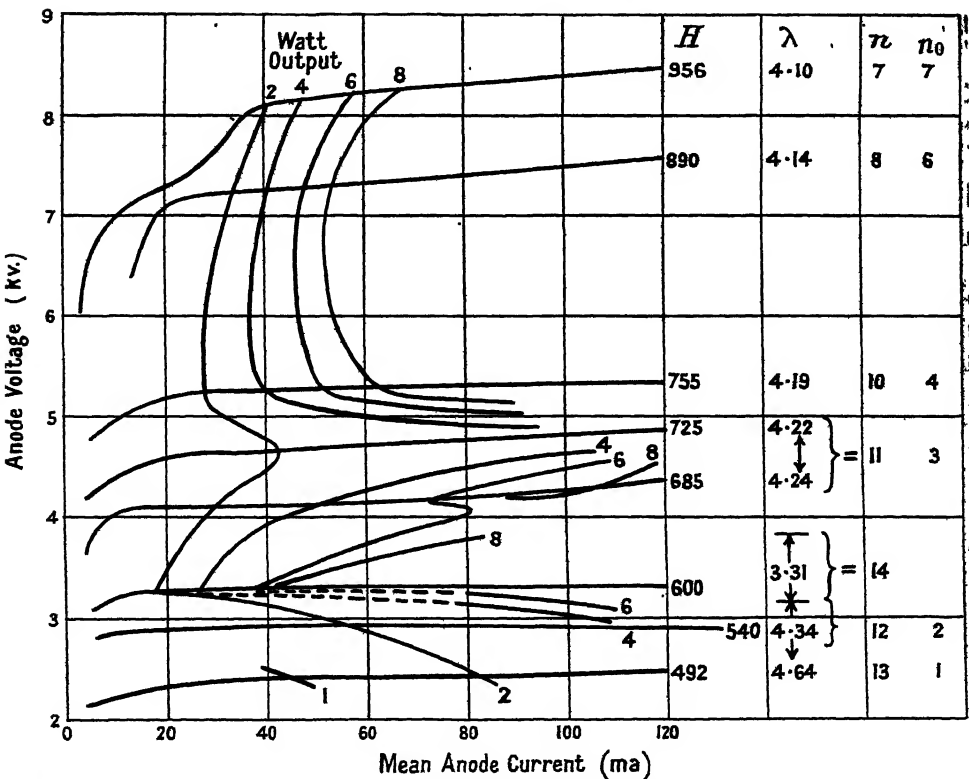


Figure 2. Variations of anode current and mode of operation with anode voltage and magnetic field strength. (H =magnetic field strength in oersteds, λ =wavelength in cm., n_0 =fundamental mode number, n =wave component mode number.)

times a wave is repeated round the circumference of the anode (Willshaw *et al.* 1946). The mode number n_0 shown in the Figure is the fundamental mode of which n is a component. It can be seen that the mode of oscillation depends entirely in the magnetic field, and oscillations are maintained in a given mode over a range of magnetic field of the order of 8%, a maximum efficiency of 6%

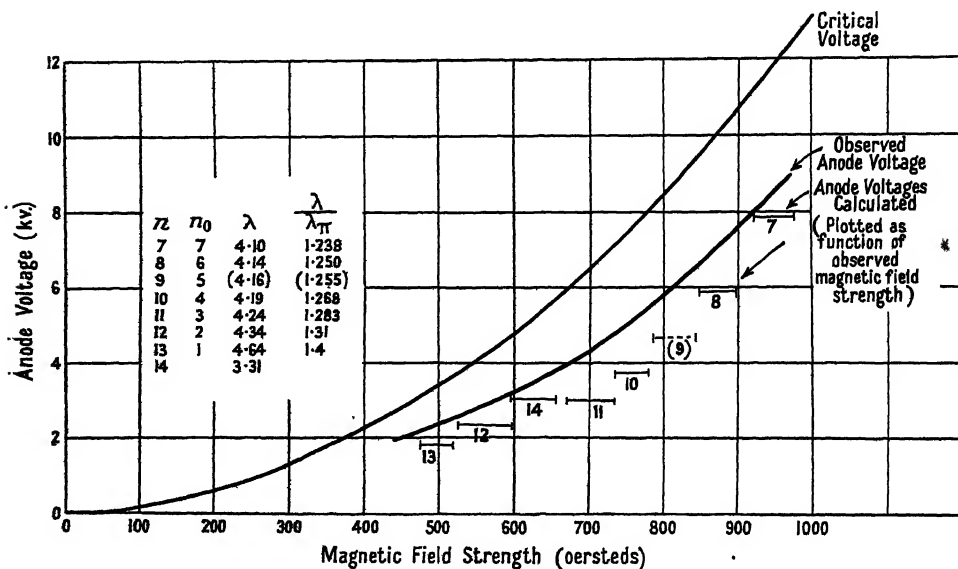


Figure 3. Relation between the observed and calculated anode voltage and magnetic field strength.

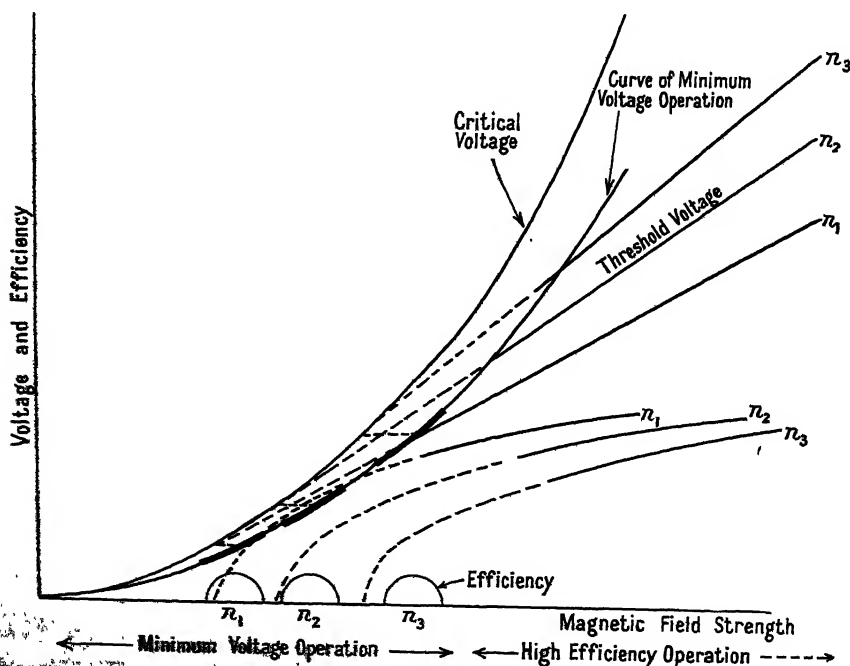


Figure 4. Relation between voltage, efficiency and magnetic field strength for minimum voltage and high efficiency operation.

being obtained in the mode $n=14$. The electronic efficiency in this mode, estimated from the ratio between power dissipated in the load and in the valve circuit, is 12%. In contrast to this, when operating as a higher efficiency pulsed valve, at an anode voltage of about 15 kv., oscillations are obtained in the mode $n=14$ (the π mode) over a wide range of magnetic field strength, as is usual for valves operated in this regime. Figure 3 shows the observed operating voltages plotted as a function of magnetic field, together with the values calculated for the different modes of operation on the hypothesis of the minimum voltage regime, assuming that the outside of the electron cloud reaches the surface of the anode so that $r_m = a$. These calculated voltages are plotted as a function of the magnetic fields at which oscillations are observed and are rather lower than the observed voltages. The critical voltages are also indicated and it is seen that operation takes place at voltages which are roughly 70% of these. Finally Figure 4 shows diagrammatically the relationship between the various parameters for the high efficiency and minimum voltage regimes of operation. The dotted extensions of the curves of high efficiency represent the values calculated from the simple expression mentioned earlier, the practical values lying much below these and falling to zero at a magnetic field greater than that at which 'minimum voltage' operation occurs.

§ 4. FIELD OF APPLICATION

It is evident from the results which have so far been achieved that useful efficiency may be obtained with a multiple circuit magnetron operating like a travelling wave tube in the minimum voltage regime. Values of anode voltage and magnetic field are several times smaller than for the high efficiency regime. The size of the valve for a given voltage is thus a maximum, and this fact and the low magnetic field involved make this mode of operation of extreme importance for work at the shortest wavelengths.

REFERENCES

- HOLLENBERG, A. V., KROLL, N., and MILLMAN, S., 1948, *J. Appl. Phys.*, **19**, 624.
 KOMPFFNER, R., 1947, *Proc. Inst. Radio Engrs.*, N.Y., **35**, 124.
 WASSERMANN, I. I., 1948, *J. Tech. Phys.*, USSR, **18**, 785.
 WILLSHAW, W. E., and RUSHFORTH, L., 1946, *J. Instn. Elect. Engrs.*, **93**, IIIA, 180.
 WILLSHAW, W. E., *et al.*, 1946, *J. Instn. Elect. Engrs.*, **93**, IIIA, 985.

Measurements of the Reflection Coefficient of Water at a Wavelength of 8.7 mm.

By D. G. KIELY

Royal Naval Scientific Service

*Communicated by L. H. Bainbridge-Bell ; MS. received 25th April 1949,
and in amended form 3rd August 1949*

ABSTRACT. An account is given of measurements (carried out in March 1947) of the reflection coefficient of water at a wavelength of 8.7 mm. over a range of angles of incidence.

The method employed is to measure the relative field strengths of the direct waves and waves reflected from a trough of water using free-space propagation and high-gain aerials.

The following electrical constants of water have been computed from the measured results (water temperature 11.1°C.): refractive index 4.40 ± 0.24 , dielectric constant 10.86 ± 2.21 , absorption coefficient 2.91 ± 0.06 , conductivity/frequency 12.82 ± 0.42 , Brewster angle 79° .

§ 1. INTRODUCTION

THIS note describes some experimental work carried out at a wavelength of 8.7 mm. to determine the reflection coefficient, conductivity, dielectric constant and absorption coefficient of water. The reflection coefficient was measured for different angles of incidence, and from these results the other constants were calculated. At this wavelength the problem of measuring the reflection coefficient was considered more optical than radio in nature and the method employed was to measure the relative field strengths of direct and reflected free-space waves using high-gain (narrow beam width) aerials and a trough of water.

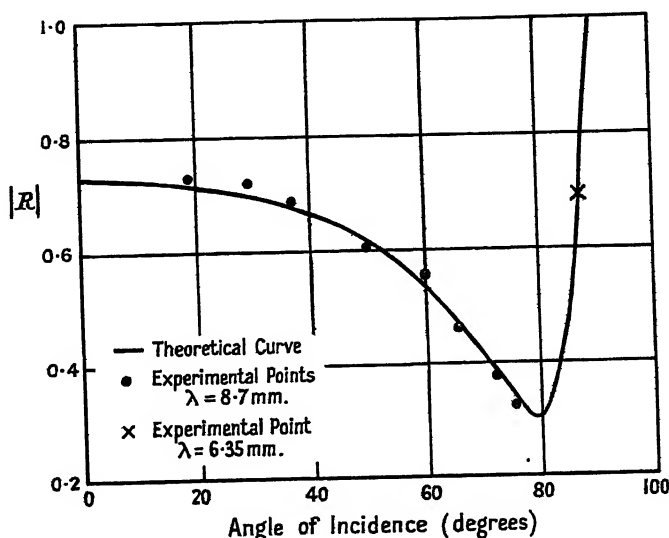
§ 2. DESCRIPTION OF APPARATUS

The experimental apparatus consisted of a transmitter and receiver, mounted on metal racks, which allowed the heights of the aerials to be varied, above a trough of water placed between them. The horizontal distance between the receiver and the transmitter was maintained constant and the total variation in height of either aerial above the water surface was approximately 4 feet. The trough dimensions were 8 ft. 9 in. by 3 ft. The aerials and radio-frequency circuits were mounted on tilting platforms attached to the racks; this allowed them to be pointed down at the water surface or directly across it. The beam width was sufficiently small for the direct and reflected rays to be received independently. Polarization was vertical.

§ 3. EXPERIMENTAL PROCEDURE

The horizontal distance between the receiving and the transmitting aerials was kept fixed. The heights of the two aerials were adjusted so that for each angle of incidence the aerials were directed at the centre of the trough. The experimental procedure was extremely simple and straightforward. The power was switched on, 30 minutes was allowed for the klystrons to become warm and stable in frequency, the transmitter and receiver were tuned, and the receiver gain was turned to a suitable level. For each angle of incidence a direct transmission across the trough

was received, and the receiver output, V_1 , and the transmitter power monitor reading noted. The aeriels were then set to the required angle of incidence, the reflected wave received, receiver output V_2 noted and the transmitter power monitor reading checked. This was repeated several times for each angle of incidence and the average value of V_1 and V_2 obtained. Only very small variations in V_1 and V_2 were noted in these repetitions. A correction was made in the value of V_2 for the difference in path length for direct and reflected waves using the inverse-distance law to give V_3 ; the ratio V_3/V_1 was then equal to $|R' + iR''|$, the modulus of the reflection coefficient, R' and iR'' being the real and imaginary components. The water temperature during the measurements was 11.1°C . The experimental results are shown in the Figure as a function of angle of incidence. The point



at 87.5° was obtained at a wavelength of 6.35 mm. by reflections from the sea during another experiment. This point has not been used in the following theoretical work.

An experiment was carried out to check the validity of the inverse-distance law of propagation for the distances involved in the above work. This experiment showed that the inverse-distance law was valid for the six measurements from 37.5° to 76° , but was not strictly valid for the two measurements at 19° and 30° incidence. These two observations have not been used in the following calculations.

§ 4. THEORETICAL ANALYSIS OF RESULTS

The symbols used are as follows: μ^2 = complex dielectric constant of water, n = real part of refractive index of water, ϵ = real part of dielectric constant of water, κ = absorption coefficient of water, f = frequency of radiation, $R = R' + iR''$ = complex coefficient of reflection of water, θ = angle of incidence, ϕ = complex angle of refraction, σ = conductivity of water.

It is suspected that $\epsilon \gg 1$ and $\sigma/f \gg 1$. Now

$$\epsilon - i \frac{2\sigma}{f} = \mu^2 = (n - i\kappa)^2. \quad \dots\dots(1)$$

If $\sin \phi = \mu^{-1} \sin \theta$, the reflection coefficient is

$$|R| = \left| \frac{\mu \cos \theta - \cos \phi}{\mu \cos \theta + \cos \phi} \right|. \quad \dots\dots (2a)$$

Now, as there is *a priori* reason to suspect that $|\mu^2| = |\epsilon - i(2\sigma/f)| \gg 1$, and since $\sin \theta < 1$,

$$\cos \phi = \sqrt{\left(1 - \frac{\sin^2 \theta}{\mu^2}\right)} = 1 - \frac{1}{2} \frac{\sin^2 \theta}{\mu^2} + \dots \quad \dots\dots (2b)$$

and the error in adopting this approximation will be at the most of the order of 2%. Hence

$$|R^2| = \frac{|\mu^2| \cos^2 \theta + 1 - 2n \cos \theta}{|\mu^2| \cos^2 \theta + 1 + 2n \cos \theta}.$$

Let
$$y = \frac{1 + |R^2|}{1 - |R^2|} = \frac{|\mu^2| \cos^2 \theta + 1}{2n \cos \theta}, \quad \dots\dots (3)$$

i.e.
$$y = ax + b/x, \quad \dots\dots (3a)$$

where $x = \cos \theta$; $a = |\mu^2|/2n$; $b = 1/2n$. In order to determine the best values of a and b , the method of least squares was adopted with suitable weighting factors. This gives the values $a = 3.163 \pm 0.065$; $b = 0.114 \pm 0.006$.

Having found a and b , the physical constants for water were obtained as follows:

$$\begin{aligned} n &= 1/2b &= 4.40 \pm 0.24, \\ \epsilon &= 1/2b^2 - a/b &= 10.86 \pm 2.21, \\ \kappa &= \sqrt{a/b - 1/4b^2} &= 2.91 \pm 0.06, \\ \sigma/f &= (1/2b^2) \sqrt{4b - \frac{1}{4}} &= 12.82 \pm 0.42. \end{aligned}$$

Using the values of a and b , a curve of $|R|$ against θ was computed from (3a). This is shown in the Figure by the experimental points. The Brewster angle, obtained by differentiation of (3a), to give the minimum value of $|R|$ is found to be 79° , as indicated in the curve. The experimental point at 87.5° incidence is in good agreement with the theoretical curve, which indicates that there is possibly little difference between the constants governing reflection from water at 8.7 and 6.35 mm.

Saxton (1947) gives theoretical curves of the variation of absorption coefficient and refractive index of water with wavelength and temperature. Comparing the above experimental results with these curves it will be seen that the agreement is quite good when the temperature difference is taken into consideration.

ACKNOWLEDGMENTS

This work was carried out for the Royal Naval Scientific Service, and permission to publish this paper is gratefully acknowledged.

The author wishes to acknowledge the assistance of his colleagues with the theoretical and experimental work carried out for the production of this paper.

REFERENCE

Saxton, J. A., 1947, *Meteorological Factors in Radio-Wave Propagation* (London: Physical Society), p. 278.

Coupling of the Ordinary and Extraordinary Rays in the Ionosphere

By T. L. ECKERSLEY *

*Communicated by G. I. Finch; MS. received 8th March 1949,
in amended form 23rd June 1949*

ABSTRACT. In this paper an approximate solution to the wave equation is given for propagation in an ionosphere in which the gradient of the density N is in the vertical, z , direction only, and in which account is taken of the earth's magnetic field. It corresponds exactly to the ray theory and expresses a quantity Z , which is the z derivative of the phase function S , by a quartic equation. Z can be represented as a function of ζ (which is proportional to N) on a four-sheeted Riemann surface, and the branch points are studied for the case of vertical incidence for which Z becomes the refractive index. By considering the branch points in the complex ζ plane, the amount of the coupling between the ordinary and the isolated extraordinary branches of the (Z, ζ) curves can be expressed as a function of the obliquity of the magnetic field. The triple splitting of rays reflected from the ionosphere, observable where the field is nearly vertical, can thus be explained, and the theory is substantiated by the observation that the polarizations of the echoes on the (P', f) records are ordinary, ordinary and extraordinary in order of increasing critical frequencies, as given by the branches of the (Z, ζ) curves.

THE problem of how a wave travels in the ionosphere is complicated by the effect of the earth's magnetic field, but a very good approximate theory can be obtained. It is the more accurate the more uniform the layer is, and it corresponds exactly to the ray theory.

By the use of Maxwell's equation we get the original wave equation in the vector form

$$\frac{1}{c^2} \frac{d^2 E}{dt^2} + \nabla^2 E - \nabla \operatorname{div} E = \left(\frac{2\pi}{\lambda} \right)^2 \frac{\nu_0'^2}{\nu_H^2 - \nu'^2} \left[\frac{\nu_H^2}{\nu'^2} \frac{H(HE)}{H^2} - \frac{i\nu_H}{\nu'} \frac{[HE]}{H} \right],$$

where $\nu_0'^2 = \nu_0^2(1 + i\alpha)$, in which ν_0 is the usual critical frequency of the medium, $\alpha = \nu_0/2\pi\nu$ the absorption term, ν_0 the collisional frequency, ν the wave frequency $\nu' = \nu(1 + i\alpha)$, and $\nu_H = eH/2\pi m$, and is the resonant frequency of the electron in the magnetic field H .

We now assume that the solution of the wave equation can be represented in the form of Jeffrey's approximate solution,

$$E = \frac{1}{(dS/dx)^{\frac{1}{2}}} E_{1,2,3} \exp(2\pi i S - 2\pi i \nu t),$$

where E_1, E_2, E_3 are the three components of E , and $S = \text{const.}$ is the phase surface, and make the approximation that the $\nabla^2 S$ term can be neglected in comparison with the $(dS/dx)^2$ terms, which implies that the gradient of the refractive index μ is everywhere small.

On reduction, we obtain three equations that can be represented in matrix form by

$$\begin{vmatrix} 1 - m^2 - Z^2 + \alpha_{11} & \alpha_{21} & \alpha_{31} \\ \alpha_{12} & 1 - Z^2 + \alpha_{22} & mZ + \alpha_{32} \\ \alpha_{13} & mZ + \alpha_{23} & 1 - m^2 + \alpha_{33} \end{vmatrix} \times E = 0. \quad \dots (1)$$

* Private address—Weatheroak, Danbury, Essex.

Here we have assumed that the gradient of density in the ionosphere is purely in the vertical z direction, so that we make $\lambda(dS/dx)=l$ and $\lambda(dS/dy)=m$, where l and m give the initial direction cosines of the wave, and $\lambda(dS/dz)=Z$. $\alpha_{11} \dots$ are given by

$$\begin{aligned}\alpha_{11} &= \xi \rho_1^2 + \zeta & \alpha_{12} &= \xi \rho_1 \rho_2 + i\eta \rho_3 & \alpha_{21} &= \xi \rho_1 \rho_2 - i\eta \rho_3 \\ \alpha_{22} &= \xi \rho_2^2 + \zeta & \alpha_{13} &= \xi \rho_1 \rho_3 + i\eta \rho_2 & \alpha_{31} &= \xi \rho_1 \rho_3 - i\eta \rho_2 \\ \alpha_{33} &= \xi \rho_3^2 + \zeta & \alpha_{23} &= \xi \rho_2 \rho_3 + i\eta \rho_1 & \alpha_{32} &= \xi \rho_2 \rho_3 - i\eta \rho_1\end{aligned}$$

in which

$$\rho_{1,2,3} = H_{1,2,3}/H,$$

where H_1, H_2, H_3 are the components of H , and

$$\xi = -\frac{\nu_0'^2}{\nu'^2 - \nu_H^2} \frac{\nu_H^2}{\nu'^2}; \quad \eta = \frac{\nu_0'^2}{\nu'^2 - \nu_H^2} \frac{\nu_H}{\nu'}; \quad \zeta = \frac{\nu_0'^2}{\nu'^2 - \nu_H^2}.$$

From the matrix we obtain an equation for Z which is biquadratic. We can thus represent Z on a four-sheeted Riemann surface, and this is sufficiently accurate except near the branch, or reflection points.

We can express the equation in terms of ζ and τ , where ζ is now defined by $\zeta = \nu_0^2/\nu^2$. This is the previous ζ with ν_H neglected, and $\tau = \nu_H/\nu$, which may be resolved into the components τ_x, τ_y, τ_z .

If we know ζ and the components of τ , we can express Z as a function of its position on the Riemann surface (Forsyth 1900).

We presume ζ and the components of τ to be known, because we cannot attack a problem without knowing the density N in the ionosphere, so that the topography of the surface is known and does not depend upon what we represent on it. All the branch points and reflection points are known and depend on ζ .

The behaviour of propagation can be represented as the behaviour of Z , and the four-sheeted Riemann surface gives everything that we require. Thus the transmission of electric waves in an ionized medium with a magnetic field present can be expressed most suitably in terms of this four-sheeted biquadratic Riemann surface. This, though not perfectly accurate, is very nearly so, except in the neighbourhood of the branch points.

For the particular case of vertical incidence, the branch points can be determined from the formula

$$Z^2 = \lambda^2 \left(\frac{dS}{dz} \right)^2 = \frac{(1-\zeta)(\gamma-\zeta) - \frac{1}{2}\zeta\tau_x^2 \pm \zeta[\frac{1}{2}\tau_x^4 + \tau_z^2(1-\zeta)^2]^{\frac{1}{2}}}{\gamma(1-\zeta) - \zeta\tau_x^2}. \quad \dots\dots(2)$$

where $\gamma = 1 - \tau^2$ and $\tau_y = 0$.

Jeffrey's solution, which gives the waves on the earth, depends upon S , and S is given by the above formula, derived from matrix (1).

Jeffrey's solution is

$$E = \frac{A}{\lambda(dS/dz)^{\frac{1}{2}}} \exp \left(\frac{2\pi i}{\lambda} S_z - 2\pi i \nu t \right), \quad \dots\dots(3)$$

which depends almost entirely on the phase S .

For vertical transmission, dS/dz can be expressed in the above form. It is biquadratic, and represents the variation of the phase of the wave. This formula gives the refractive index and all the branch points of S , which are either zeros or infinities of the above expression.

We find that the group velocity is zero at all these points, which means that a pulse will come up to the branch points, stop, and then go down again. These branch points are therefore the reflection points. We give an example in Figure 1, which shows the refractive index $\lambda(dS/dz)$ as a function of the quantity ζ , which is the ratio of the frequencies ν_0^2/ν^2 , where $\nu_0^2 = Ne^2/\pi mc^2$.

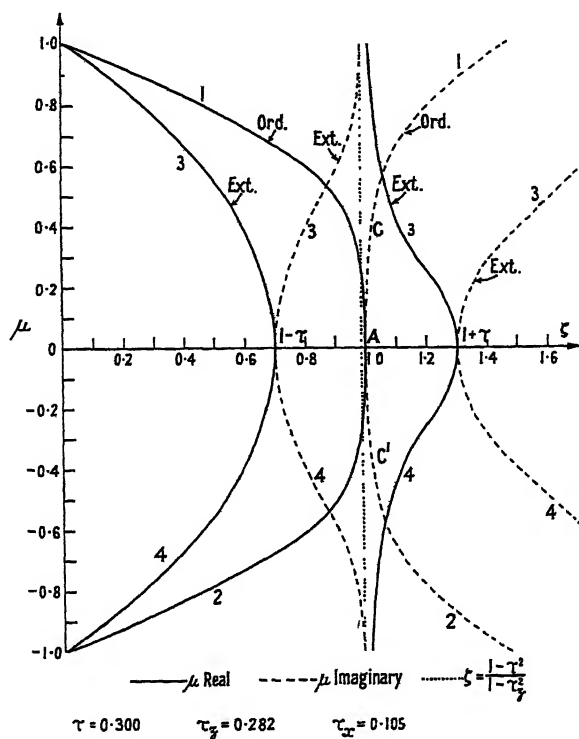


Figure 1. Refractive index as a function of $\zeta(\propto N)$, showing reflection at the branch points.

We give here the four branch points in terms of ν_0 :

1. $\nu_1 = \nu_0/(1-\tau)^{\frac{1}{2}}$ Extraordinary
2. $\nu_2^* = \{(1-\tau_z^2)/(1-\tau^2)\}^{\frac{1}{2}}\nu_0$
3. $\nu_3 = \nu_0$
4. $\nu_4 = \nu_0/(1+\tau)^{\frac{1}{2}}$

approximately, since the τ 's are functions of ν_1, ν_2 etc.

More explicitly, they can be expressed in the form

1. $\nu_1^2 - \nu_1 \nu_H = \nu_0^2,$
2. $\nu_2^2 \{(\nu_2^2 - \nu_H^2)/[\nu_2^2 - (\nu_H^2)^2]\} = \nu_0^2,$
3. $\nu_3^2 = \nu_0^2,$
4. $\nu_4^2 + \nu_4 \nu_H = \nu_0^2.$

The branch or reflection points are shown in Figure 2. This gives the nature of the four-sheeted Riemann surface. The branch points correspond to the normal ordinary and extraordinary rays.

One thing is clear. The ordinary and extraordinary rays are on different sheets of the Riemann surface. When the transmission is vertical and the magnetic field is completely vertical, the sheets are in two unconnected pairs, and there is no means of passing from an ordinary sheet to an extraordinary one or vice versa. There is no coupling between the rays. This is discussed in fuller detail later, but this outline gives an idea of the nature of the coupling between the two.

We can now go on to consider the Riemann surface method of representation. The results contained in the following pages give us an understanding of the coupling process. No criterion of the relative probability of appearance or the

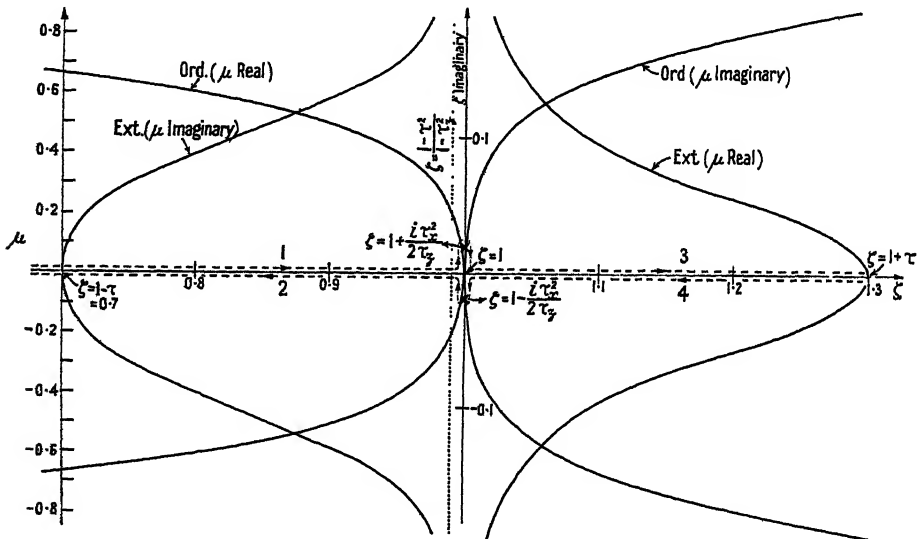


Figure 2. Refractive index as a function of $\zeta(\propto N)$, showing branch points in the complex ζ plane.

degree of intensity of the rays or pulses of these critical frequencies has ever been given, and the following method of analysis gives definite conclusions with regard to this point. The original analysis, giving critical frequencies alone, is like the older phase integral quantum theory, in which the various spectral lines were given (corresponding to the critical frequencies) without any indication of their intensities.

We have represented the waves on a four-sheeted Riemann surface on which the branch (or reflection) points will be seen. It has been pointed out that the phase and group time increase continually up to and through the branch points and follow on the downward branch to $\zeta = 0$, giving the complete phase and group delay of a wave or impulse travelling along the upward branch (dS/dz) positive to the reflection point ($dS/dz = 0$) (generally but not necessarily) and down the lower branch (dS/dz) negative to the starting point.

The conditions of reflection can therefore be best studied by examining the branch points and branch lines of the four-sheeted Riemann surface on which (dS/dz) is most suitably represented.

This surface, for the case discussed above, is shown in Figure 2, where the surface represents the complex ζ plane. The study of expression (2) shows that branch points occur at

$$\begin{aligned} \zeta = 1 - \tau \text{ (one),} & \quad \zeta = 1 \text{ (one),} & \quad \zeta = 1 + (i\tau_x^2/2\tau_z) \text{ (two),} \\ \zeta = 1 - (i\tau_x^2/2\tau_z) \text{ (two),} & \quad \zeta = 1 + \tau \text{ (one),} & \quad \zeta = \infty \text{ (unless } \zeta = z^2), \\ & \text{and} & \quad \zeta = \gamma/(1 - \tau_z^2) \text{ (one),} \end{aligned}$$

when the refractive index becomes infinite.

If ζ negative is cut off by a perfectly conducting plane, there are two further branch points at $\zeta = 0$ representing the complete reflection of the ordinary and extraordinary rays at the surface. The branch lines ending at the branch points should be noted.

Any path which crosses a branch line goes from one sheet, say 1, to another sheet, say 2. The branch line ends at any branch point where the two are interchanged.

The analysis of the four-valued expression for μ in this manner reveals significant features. When the incidence is vertical and the earth's magnetic field is also vertical, then there are two independent solutions on two independent pairs of sheets, i.e. the 1-2 pair and the 3-4 pair, and there is no interchange between the pairs.

When, however, the magnetic field is at all oblique, both the longitudinal and the transverse components of H are present. All four sheets are then linked together. There are branch points where 1 and 3 and where 2 and 4 are interchanged, and it is always possible to choose a path which goes from any one sheet to any other. The ordinary and extraordinary rays are then coupled together, as well as the ordinary and isolated rays, although usually to a very slight extent.

This analysis may be used for estimating or computing the intensities of the echoes corresponding to the four critical frequencies already given, and for calculating the coupling between the ordinary and extraordinary rays.

In the first place it should be noted that it will be very improbable that the ray with critical frequency ν_z (pointed out by Rai 1937) will ever be obtained with observable intensity, for the path of the extraordinary echo which is reflected at $\zeta = \gamma/(1 - \tau_z^2)$ (see Figure 2), corresponding to the critical frequency ν^2 , has to traverse the region from $\zeta = 1 - \tau$ to $\zeta = \gamma/(1 - \tau_z^2)$, where the refractive index is large and entirely imaginary, before being reflected.

There is, however, a much more likely case, in which, instead of the normal ordinary and extraordinary doublet, which separates as a split echo near the critical frequency, there is a triplet ordinary, ordinary, extraordinary: o., o., ex. (see Figures 2, 3 and 4).

The three critical frequencies here shown are given by

$$\begin{array}{ccc} \zeta = (1 + \tau), & \zeta = 1 & \text{and} \quad \zeta = (1 - \tau). \\ \text{o.} & \text{o.} & \text{ex.} \end{array}$$

Since $\zeta = \nu_0^2/\nu^2$, where ν is the critical frequency, the three critical frequencies are

$$\begin{array}{ccc} \nu_0/(1 + \tau)^{\frac{1}{2}}, & \nu_0 & \text{and} \quad \nu_0/(1 - \tau)^{\frac{1}{2}} \\ \text{o.} & \text{o.} & \text{ex.} \end{array}$$

where $\nu_0^2 = Ne^2/\pi mc^2$.

The last two are the normal critical frequencies and the first an extra one.

This split into three branches was first observed by Harang (1936), who showed the three branches (a), (b) and (c) each separated by a frequency interval $\nu_H/2$ appropriate to the critical frequencies given above.

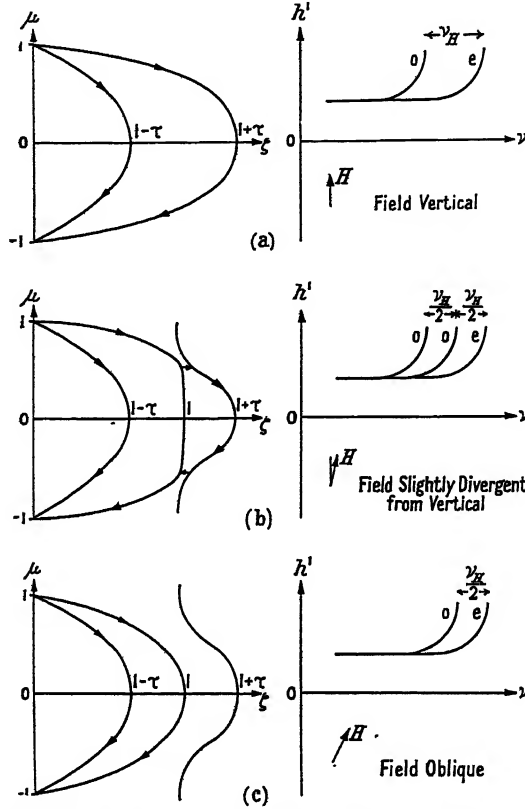


Figure 3. Similar to Figure 1, showing the corresponding $\bar{h}'(\nu)$ curves with the production of a triplet.

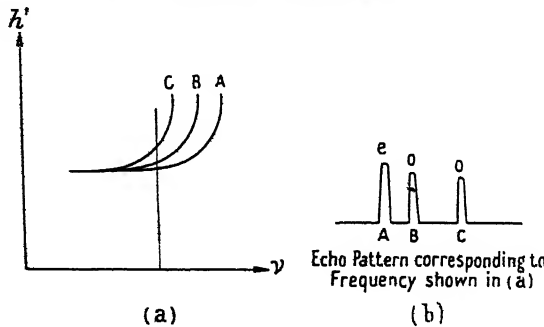


Figure 4. (h', ν) curve, showing the triplet with the corresponding echo pattern and polarizations.

The mechanism of this is illustrated in Figures 1, 2 and 3. Turning first to Figure 1, it will be seen that if we could bridge the gap from the ordinary ray (1) to the isolated ray (3) at C, and then move along the 3-4 path through the apex at D to C', bridging the gap again at C' on to the downward ordinary branch 2, we should form a path the critical frequency of which would be $\nu_0/(1+\tau)^{1/2}$,

corresponding to the added ordinary ray observed by Harang. This ray was also observed by the author on the cathode-ray oscillograph in 1933, the order of delay being ex., o., o. (Eckersley 1933, also Meek 1948, Newstead 1948).

The polarization of this extra ray, determined as it is by the conditions when leaving the ionosphere on the downward ordinary path, will obviously correspond to that of the ordinary ray.

That the physics of such a transmission is possible will be seen by proceeding to the limit where $\tau_x \rightarrow 0$ and H , the magnetic field, is purely vertical. When τ_x is very small, but H not quite vertical, the analysis shows that the isolated branch 3-4 very nearly touches the ordinary ray branch at CC' , and in the limit when $\tau_x \rightarrow 0$, the two actually join, and the paths CAC' and $C\infty$ disappear. We are then left with the two simple rays with critical values of ζ , viz. $1 - \tau$ and $1 + \tau$. These are the extreme rays of the triplet, and have a frequency difference ν_H instead of the usually observed difference $\nu_H/2$.

It is hardly conceivable, physically, that this process should occur discontinuously at $\tau_x = 0$, and we should expect, on physical grounds, to find all three branches in the neighbourhood of $\tau_x = 0$ for the reason that the isolated branch 3-4 is closely coupled to the ordinary branch 1-2 when τ_x is small.

We can express these ideas in a more exact form by considering the Riemann surface again. We have to find the total attenuation on integrating this along a path which, starting in sheet 1, passes continuously to sheet 3 round the 1-3 branch point, again passes continuously from sheet 3 to sheet 4 round the 3-4 branch point at $\zeta = 1 + \tau$, and then from sheet 4 to sheet 2 round the 4-2 branch point and back to earth. In sheet 2, as in sheet 1, N , the number of electrons, is zero at $\zeta = 0$.

This represents, in a more exact mathematical form, the physical path previously discussed. Thus it passes, in sheet 1, as far as $\zeta = 1$; then it goes from 1 to $1 + i(\tau_x^2/2\tau_x)$, where, ζ being complex, there is an attenuation. After passing round the branch point 1-3, it reaches sheet 3, and ζ decreases on this sheet from $1 + i(\tau_x^2/2\tau_x)$ to the neighbourhood of $\zeta = 1$, again making a contribution to the attenuation. We then integrate along the path from $\zeta = 1$ to $\zeta = 1 + \tau$ in sheet 3, where the path is along the real axis and the integral is real, so that there is no contribution to the attenuation. At this point, $1 + \tau$, it goes from sheet 3 to sheet 4.

Finally, we require to pass from sheet 4 to sheet 2 to continue the path down to earth, and to do so we have again to introduce attenuation by rounding the branch point at $\zeta = 1 - i(\tau_x^2/2\tau_x)$ where ζ and the integrand are again both complex. We then reach branch 2, the ordinary ray, at $\zeta = 1$ and proceed, without attenuation, along this path to earth again at $\zeta = 0$.

The imaginary part of μ is definitely specified at each point of the path, so that the attenuation can be uniquely determined.

It will be observed that the path and integrand are only complex in the regions between $\zeta = 1$ and $\zeta = 1 + i(\tau_x^2/2\tau_x)$ and between $\zeta = 1$ and $\zeta = 1 - i(\tau_x^2/2\tau_x)$, and that this path tends to zero as $\tau_x \rightarrow 0$. Therefore the total attenuation of this spurious ordinary echo decreases as $\tau_x \rightarrow 0$. This was also obvious in the previous more definitely physical treatments.

Another point must be considered: The integral should be taken with respect to z , i.e. the distance, but as expressed on the Riemann surface, μ , or $\lambda(dS/dz)$, is a function of ζ only.

The attenuation or imaginary part,

$$\int \lambda \frac{dS}{dz} dz = \int f(\zeta) dz,$$

can be expressed in the form $\int f(\zeta) d\zeta(dz/d\zeta)$.

Now the first part of this integral is a quantity only depending on τ_x, τ_z, τ_y , not on the distribution of density in the layer. The whole integral is small when the gradient of density in the layer is large, and vice versa.

Therefore the attenuation equals $\int f(\zeta) \frac{d\zeta}{d\zeta/dz}$, and is likely to be small when $d\zeta/dz$, the ionic gradient, is large, and vice versa.

It follows that the conditions in which the triplet form with critical frequencies $\nu_0/(1-\tau)^{\frac{1}{2}}, \nu_0, \nu_0/(1+\tau)^{\frac{1}{2}}$ is likely to appear are: (i) in northern latitudes where τ_x is small, and (ii) in regions where the gradient of N with respect to z is large. It is significant that this triplet form has only been observed in high latitudes.

Cases which may well represent this type of transmission were noted in 1933 (Eckersley 1933). On such occasions well defined triplets were observed on frequencies in the neighbourhood of the critical frequency. The polarization observed in the cathode-ray indicator was invariably as shown in Figure 4.

A and B are the normal extraordinary and ordinary pair, the extra echo C being of ordinary polarization.

These three correspond to the branches with critical frequencies:

A	B		C
$\nu_0/(1-\tau)^{\frac{1}{2}},$	ν_0	and	$\nu_0/(1+\tau)^{\frac{1}{2}}$ (the extra echo).
ex.	o.		o.

We may use the curves in Figure 1 in the way just described to calculate the coupling coefficient between the ordinary and extraordinary rays. Thus, considering an up-going ray in the ordinary 1-2 sheet, we may bridge the gap, as before, to the isolated ray in the 3-4 sheet. We then pass to the branch point at

$$\zeta = \gamma/(1-\tau_x)^{\frac{1}{2}}, \quad dS/dz = \infty$$

and back along the imaginary branch in the 3-4 sheet to D, and finish along the real part of the extraordinary branch which goes down to earth again. It is clear that the total attenuation is large and, therefore, that the coupling will be small. The contribution to the attenuation is especially large between S and D. The same remarks as regards the gradient of ζ with respect to z apply equally well here. Thus if $d\zeta/dz$ is large enough, the total attenuation along the path may be made as small as we please.

The truth of this statement may readily be seen by proceeding to the limit where $d\zeta/dz$ becomes so large at one point that we may consider this to be the surface of separation. This case may easily be worked out according to the full wave theory. Without going into detail, it may be said that in medium 1 (see Figure 5) there may be two characteristic waves, o_1 and ex_1 . Let the polarization ratio X/Y be given by $P_{o,1}, P_{ex,1}$.

Therefore, in medium 2, we shall have $P_{o,2}, P_{ex,2}$.

Suppose a wave of the type o is sent through the first medium. Since $P_{o,2}$ and $P_{ex,2}$ are not the same as $P_{o,1}$ and $P_{ex,1}$, and since there is an appreciable

z component of the electric force, the reflected wave will not be only of the o. type, but will consist of o. and ex. in suitable proportions to satisfy the boundary conditions.

If the path in medium 1 is sufficiently long, and if the incoming signal consists of a short wave impulse, the reflected wave will consist of an o. and an ex. impulse which, after travelling sufficiently far, will be separated into two pulses.

Thus, by sending into the medium an o_1 impulse, we get a pair, R_{o1} and R_{ex1} , reflected or refracted back into medium 1 again.

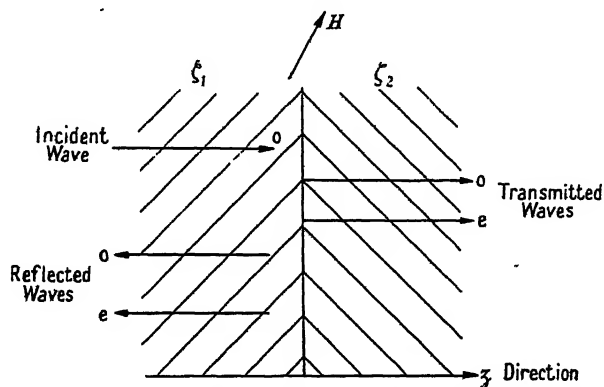


Figure 5. Illustrating the full wave solution at a surface of separation between ζ_1 and ζ_2 .

Thus there is obviously a coupling between the o_1 and ex_1 rays where the gradient is so steep that in the limit we may consider it to form a definite surface of separation.

This is an extreme illustration of the effect of coupling already referred to.

This coupling has, in general, been calculated by the relatively inaccurate four-sheeted Riemann surface method because it is very difficult to get a suitable solution of a fourth order differential equation. As reckoned by this method, the amount of coupling may be a little inaccurate, but where the media on each side of the surface of separation are uniform, it is possible to use a completely definite wave solution which agrees with the above and shows, as in the last analysis, that there is a definite coupling between the o. and ex. waves. Some German papers also show this, and I think we may rest assured that this coupling exists.

RECAPITULATIONS

1. Triple splitting of rays reflected from the ionosphere has been observed in the polar regions where the earth's magnetic field is nearly vertical (see Figure 4). Well defined triplets are almost invariably of the type o., o., ex. in terms of critical frequency. In terms of delay time they are ex., o., o. They have been observed by Harang in Norway, J. M. Meek in Canada and by the author as far south as England. They have also been observed by Newstead in Hobart, Tasmania, but as far as is known they have never been observed in any equatorial region.

2. These triplets can be explained in terms of the four-sheeted Riemann surface, which expresses the propagation in the ionosphere. They are due, as shown in Figure 3 (b), to the coupling between the isolated and ordinary rays

This coupling does not occur in equatorial regions. The Riemann surface theory, which is approximate, has been confirmed in certain cases by the wave theory, which is absolutely correct.

3. It is in virtue of a coupling between the o. and isolated rays that a triplet is formed. When τ_z is zero, and we are consequently at the pole, there is no definite isolated ray, and the two extremes of the triplets are formed with a frequency separation of ν_H approximately, and not the usual $\nu_H/2$ approximately.

4. The fact that the observed features of the triplets can be explained by the Riemann surface theory, and can be derived from it, is a very good proof that this theory is correct. These features are: (i) the triplet is of the ex. o., o. type; (ii) it is confined to high positive and negative latitudes.

Triplets were noted by the writer in England as early as 1933; experimental results were given, but the explanation was not published at the time.

REFERENCES

- ECKERSLEY, T. L., 1933, *Proc. Roy. Soc. A*, **141**, 710; 1948, *Nature, Lond.*, **161**, 597.
 FORSYTH, A., 1900, *Theory of Functions* (Cambridge: University Press).
 HARANG, L., 1936, *Terr. Mag. Atmos. Elect.*, **41**, 160.
 MEEK, J. M., 1948, *Nature, Lond.*, **161**, 597.
 NEWSTEAD, G., 1948, *Nature, Lond.*, **161**, 312.
 RAI, R. N., 1937, *Proc. Nat. Inst. Sci.*, **3**, 307-317.

LETTERS TO THE EDITOR

The Debye Effect in Electrolytes

P. Debye (1932) has predicted the existence of an alternating electric potential Φ at a point in an electrolyte through which a sound wave is passing. His equation is

$$\Phi = \frac{m_H}{\epsilon} g a_0 \frac{\sum (p_j \zeta_j M_j / \rho_j)}{\sum (p_j \zeta_j^2 / \rho_j)} \left\{ \frac{4\pi l / D \omega}{[1 + (4\pi l / D \omega)^2]^{\frac{1}{2}}} \right\}$$

where ω is the angular frequency, g the velocity of sound in solution, a_0 the particle velocity amplitude, l the conductivity of the solution, D the dielectric constant of the solvent, m_H the mass of a hydrogen ion, ϵ the electronic charge, p_j the relative number per unit volume of ions of valency ζ_j , gram ionic mass M_j and frictional coefficient ρ_j . The deduction of this result has been followed by mathematical work by which corrections for various smaller effects have been proposed.

For frequencies less than tens of megacycles per second the last factor in the equation is of the order of unity and frequency becomes unimportant. For a binary salt with univalent ions Debye's result then reduces to

$$\Phi = 1.4 \times 10^{-7} a_0 \frac{M_1 \rho_1 - M_2 \rho_2}{1/\rho_1 + 1/\rho_2} \text{ volts.}$$

With dilute solution in a given solvent, electrolysis experiments give values for ionic mobilities differing little from ion to ion or with concentration. Hence for the purpose of determining the order of magnitude of the effect, we may put $\rho_1 \simeq \rho_2$ so obtaining a result roughly proportional to $(M_1 - M_2)$. The order of Φ should therefore be about 1 microvolt when a_0 is 1 cm/sec.: a sound intensity of about 1/100 watt/cm². This led Debye to express confidence in the early detection of the effect. In fact, no experimental evidence for it in electrolytes was found until the recent work of Yeager, Bugosh, Hovorka and McCarthy (1949), although an effect of about one hundredth of the estimated order of magnitude was recorded by Rutgers (1946) for colloidal solutions.

For purposes of calculation we should use for M_1 and M_2 the effective masses of the ions, including the mass of any solvation sheath. The use of the available conflicting results for solvation numbers for calculation of the relative values of Φ for the five salts used by the above workers shows no correlation with their experimental results. In fact these results vary so little between one salt and another in water as to suggest that the solvent is largely responsible for the effective masses of the ions. That is, there is either a large solvation sheath masking the effect of the ions proper, or the equation does not take into account all the factors involved. Available results from which conclusions may be drawn are insufficient and work with selected salts and a variety of solvents is obviously very necessary. Such results should serve to obtain quantitative data, not only about the Debye effect, but also about the mechanism of solvation.

A principal experimental difficulty is the possibility of direct electromagnetic coupling between the radiofrequency source and the detecting apparatus. Such coupling would limit the sensitivity of the amplifier which could usefully be employed. Yeager, Bugosh, Hovorka and McCarthy used elaborate screening precautions to detect the effect and their method thereby made measurement of the sound intensity impracticable. However their results appear to confirm that the effect is smaller than expected from the Debye equation and for reasonable sound intensities is therefore not much above noise level.

A solution would seem to be the use of a method by which a pulse of sound is sent through the experimental liquid, the duration of the pulse being such that, with the length of sound path used, the source is inoperative during detection of the effect.

An apparatus of this kind, by which any electromagnetic pick-up is separated in time from the desired signal, is nearing completion in this laboratory. The frequency of 465 kc/s. has been chosen so that the sound wavelength in the experimental liquid may be sufficiently long for the use of probes spaced half a wavelength apart for detection of the potential variations; at the same time standard intermediate frequency transformers may be used for the amplifier. At this frequency the absorption of sound in the electrolyte will be small and a long sound path may consequently be used. This makes it possible to use a long pulse, of the order of one millisecond in duration. The amplifier required for such a pulse need have a bandwidth of only 1,000 c/s. and the noise level will therefore be low. The method is essentially a progressive wave method, reflections from the end of the tank being avoided by the use of an absorbing device. Since a large tank may be used the sound field at the detecting probes is calculable and no difficulty is expected in determining the intensity there. The pulses are to be displayed on a cathode-ray tube and compared with signals from a standard generator. It is hoped to investigate thoroughly the dependence of the effect on the various parameters involved.

Physics Department,
University College,
Leicester.
7th October 1949.

A. N. HUNTER.

DEBYE, P., 1933, *J. Chem. Phys.*, **1**, 13.

RUTGERS, A., 1946, *Nature, Lond.*, **157**, 74.

YEAGER, E., BUGOSH, J., HOVORKA, F., and MCCARTHY, J., 1949, *J. Chem. Phys.*, **17**, 411.

Electron Optics of the Three-Stage Electron Microscope

Electron microscopes employing three stages of electronic magnification are now becoming more generally used, but as yet no adequate account of their electron optics has been published. Le Poole (1947) gave some ray diagrams, which, while they are accurate, are not sufficiently comprehensive to give a clear picture of the situation.

In connection with the building of an electron microscope in collaboration with Mr. J. F. Brown, and under the direction of Professor G. I. Finch, the conditions of conjugate foci between the electron lenses were examined by Gaussian optical methods, which, while they do not give a strictly accurate quantitative analysis of the situation, do indicate the approximate positions and method of formation of the images. The analysis is reproduced in the Table, and Figures 1, 2, 3 and 4 are the ray diagrams which were deduced from the data summarized in this Table.

Condition of Conjugate Foci between Objective Lens and Intermediate Lens of a Three-Stage Electron Microscope.

u (objective lens) = 0.8 cm.; v (intermediate lens) = 17 cm.; objective-intermediate lens separation = 15 cm.; magnification of projector lens = $M_P = 120\times$.

Objective lens			Intermediate lens			$M_0 \times M_I \times M_P$	Corresponding ray diagram
v (cm.)	f (cm.)	Magn. (M_0)	f (cm.)	Magn. (M_I)	u (cm.)		
32.5	0.785	38	—	1	—	4560	2-stage
44.6	0.786	55.8	40	0.575	— 29.6	3860	Figure 3. Low magnification (a)
54.3	0.788	67.9	30	0.433	— 39.3	3550	
68.1	0.791	85	25	0.32	— 53.1	3380	
157	0.797	196.5	20	0.12	— 142	2830	
321	0.798	400	18	0.0555	— 306	2660	
2905	0.7998	3635	17.1	0.0059	— 2980	2565	Figure 4. Low magnification (b)
∞	0.8		17	$1/\infty$	∞	2555*	
— 247	0.803	309	16	0.0684	256	2530	
— 64.3	0.81	84	14	0.220	79.3	2090	
— 25.8	0.825	32.2	12	0.43	40.8	1645	
— 4.1	0.981	5.13	9	0.916	19.1	565	Figure 2. Diffraction
			8.08	1.168	15	0	
5.72	0.701	7.15	6	1.885	9.28	1000	Figure 1. High magnification
7.91	0.728	9.9	5	2.47	7.09	2920	
9.77	0.741	12.21	4	3.35	5.23	7700	

* Interpolated.

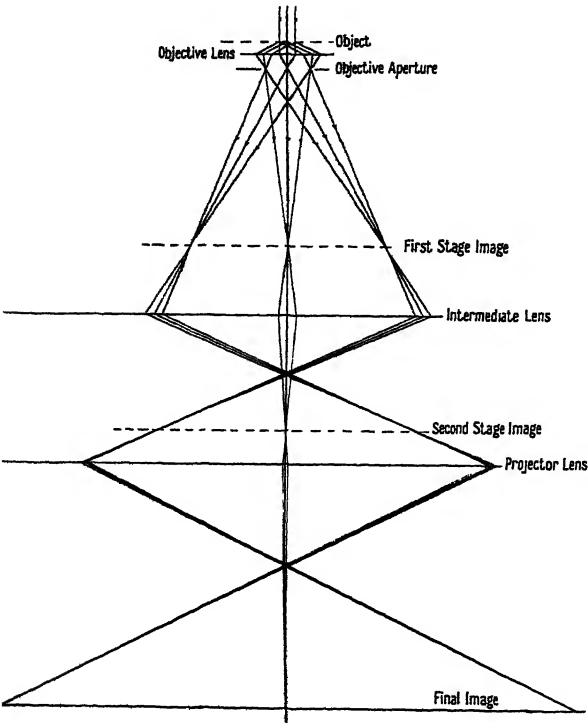


Figure 1. High magnification.

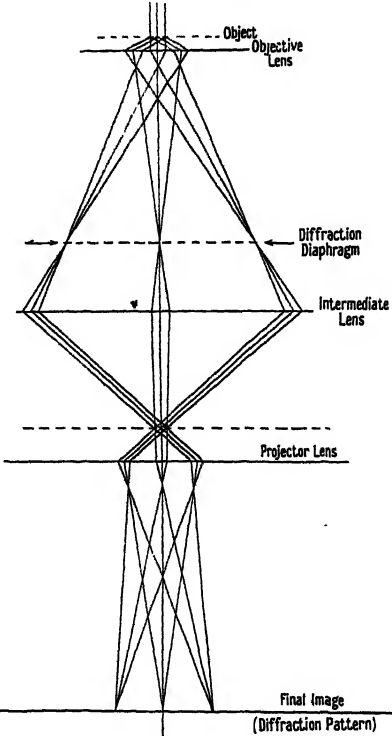


Figure 2. Diffraction.

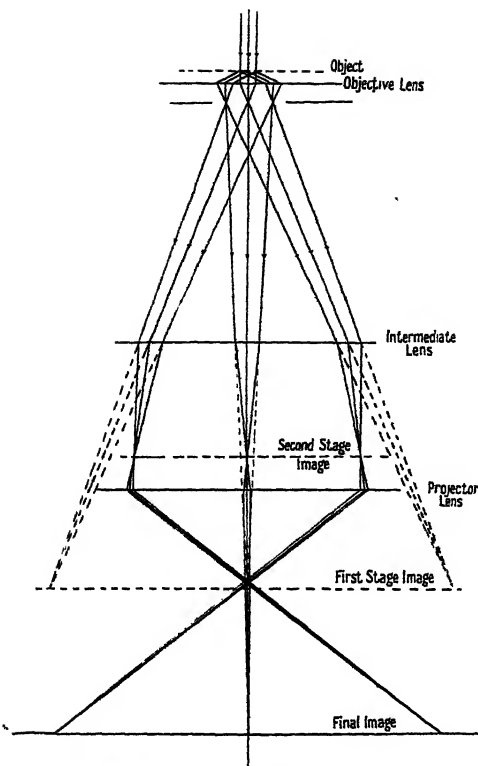


Figure 3. Low magnification (a).

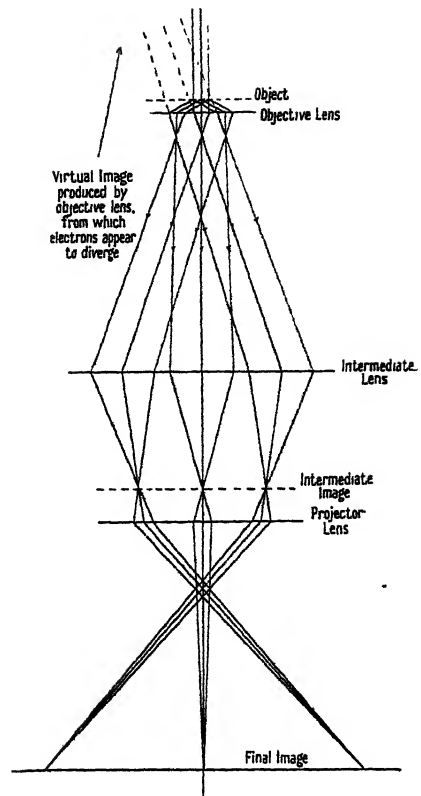


Figure 4. Low magnification (b).

Figure 1 indicates the electron paths when the microscope is producing a high magnification final image. If from this condition the power of the intermediate lens is decreased sufficiently, it is possible to obtain a zero magnification image, and if no objective aperture is present to remove the higher order diffracted rays, a diffraction pattern of the specimen is seen on the final screen (Figure 2). If the power of the intermediate lens is further decreased and the objective lens refocused, then a low magnification image is produced on the final screen, and the optics of this image are best explained in the following way.

If the intermediate lens is switched off, and the objective lens refocused, a two-stage image is formed. If now the intermediate lens is switched on at very low power, then the magnification of the image is slightly decreased; this condition is shown in Figure 3. Upon further increasing the power of the intermediate lens, a stage is reached where the objective lens no longer produces a real image; this is the case illustrated by Figure 4. Further increases in intermediate lens power produces again the diffraction pattern.

These theoretical deductions were borne out by practical measurements of focal lengths of the lenses of the electron microscope for the various conditions of image formation. It was also found that, when a low magnification was required, the image formed by using the intermediate lens in the condition corresponding to Figures 3 or 4 showed less distortion than that produced when the conditions of Figure 1 obtained.

I am indebted to Mr. J. F. Brown, with whom the work was carried out, and to Professor G. I. Finch for his helpful supervision.

Applied Physical Chemistry Laboratory,
Imperial College, London S.W. 7.*
20th September 1949.

C. E. CHALLICE.

LE POOLE, J. B., 1947, *Philips Tech. Rev.*, 9, 33.

* Now at National Institute for Medical Research, Hampstead, London N.W. 3.

REVIEWS OF BOOKS

Metal Rectifiers, by H. K. HENISCH. Pp. xi+156. First Edition. (Oxford: University Press, 1949.) 15s.

This book belongs to the series of monographs on Physics and Chemistry of Materials. I always find it difficult to define exactly the cross section of physicists or engineers to whom such a monograph is addressed. In general the practical expert will find little new in the chapters on manufacture, while he might find the chapters dealing with theory hard going because they are, of necessity, rather condensed. On the other hand the theoretical physicist often finds that the description of the practical methods is somewhat incomplete. It appears to me that in this volume the author has managed to find a happy combination of describing the essentials of practical manufacturing methods without getting lost in details, and of presenting the theory in such a way that it is not too difficult to follow for the practical engineer, though I myself would have liked to have seen the theoretical Chapter V somewhat amplified and some of the figures in it described in more detail. On the other hand, a full bibliography enables the interested reader to study the subject in greater detail.

After a short history of rectifier development the manufacture of the most important types of rectifiers is described. These comprise cuprous oxide, selenium, sulphide and germanium rectifiers.

In the third chapter characteristic rectifier properties are described. Starting with the rather complex voltage-current characteristics of the various types of rectifier, D.C. resistance and incremental or A.C. resistance are described as well as the phenomena of hysteresis, creep, ageing and the effects of temperature. The self capacitance represents an A.C. shunt and a number of equivalent networks are given together with an account of the variations of the self capacitance with the conditions of measurement such as direct voltage, time, etc. In view of the importance of "noise" in modern circuitry, the question of noise would appear to deserve more than one short paragraph at the end of the chapter.

Chapter IV deals with measurements of rectifier properties and the pitfalls that may be encountered in these.

The longest chapter in the book is, quite rightly, the one dealing with modern theories of rectification. After a discussion of the structure of semiconductors and the mechanism of electrical conduction in solids, modern concepts of insulators, intrinsic semiconductors and impurity semiconductors are given.

Next, the theories of the chemical and physical barrier layers are explained and the theoretical and experimental results compared. Chapter VI deals shortly with earlier theories of rectification.

Two short chapters on rectifier operation and future developments conclude this highly recommendable book. As far as the bibliography is concerned, replacement of the simple asterisks by numbers or by an indication of the corresponding page in the text would facilitate finding one's way back into the text when scanning through the list of references. Seven of the references mentioned in the text have not received an asterisk, and the one mentioned on page 129 has the wrong number.

H. G. LUBSZYNSKI.

CONTENTS FOR SECTION A

	PAGE
Editorial	1
Dr. K. MENDELSSOHN and Mr. J. L. OLSEN. Heat Transport in Superconductors	2
Dr. E. R. ANDREW and Mr. J. M. LOCK. The Magnetization of Superconducting Plates in Transverse Magnetic Fields	13
Prof. M. H. L. PRYCE. A Modified Perturbation Procedure for a Problem in Paramagnetism	25
Dr. R. P. PENROSE and Dr. K. W. H. STEVENS. The Paramagnetic Resonance from Nickel Fluosilicate	29
Prof. M. H. L. PRYCE and Dr. K. W. H. STEVENS. The Theory of Magnetic Resonance-Line Widths in Crystals	36
Dr. T. R. KAISER. On the Capture of Particles into Synchrotron Orbits	52
Dr. T. R. KAISER and Mr. J. L. TUCK. Experiments on Electron Capture and Phase Stability in a 14 mev. Synchrotron	67
Letters to the Editor :	
Dr. W. EHRENBERG. The Electric Conductivity of Simple Semiconductors	75
Reviews of Books	77
Contents for Section B	78
Abstracts for Section B	79

ABSTRACTS FOR SECTION A

Heat Transport in Superconductors, by K. MENDELSSOHN and J. L. OLSEN.

ABSTRACT. The heat conduction processes in superconductors have been discussed on the basis of the analogy with liquid helium II. It is suggested that under certain conditions the superconductive metal can exhibit a type of heat transport corresponding to the heat flow associated with the fountain effect. The heat conductivity of a number of pure metals and alloys has been measured in the normal and in the superconductive state and the results have been analysed with reference to the hypothesis mentioned above. In addition to the change of heat conduction with temperature, the magnetic hysteresis of the heat conduction has been investigated. The use of these phenomena as make-and-break thermal contacts at very low temperatures has been suggested.

The Magnetization of Superconducting Plates in Transverse Magnetic Fields, by E. R. ANDREW and J. M. LOCK.

ABSTRACT. The magnetization curves of thin superconducting tin plates were measured in transverse magnetic fields in order to investigate the nature of the intermediate state in such specimens. The curves showed peak magnetizations much higher than $H_c/4\pi$, indicating the presence of fields much greater than the critical field at the edge of the plates. This effect is qualitatively similar to that predicted by Landau's theory of the intermediate state, but quantitative agreement is not obtained. Resistance measurements on thin strips of tin in transverse fields showed that resistance only reappears for fields considerably greater than those required to start the destruction of superconductivity.

A Modified Perturbation Procedure for a Problem in Paramagnetism, by M. H. L. PRYCE.

ABSTRACT. A modified perturbation technique is described for problems in which second-order effects are comparable in magnitude with first-order effects, where orthodox methods break down. It is applied to the energy levels of paramagnetic ions in a crystal, giving an effective Hamiltonian in which the Stark splitting, the anomalous g -value and the temperature-independent paramagnetism are clearly exhibited.

The Paramagnetic Resonance from Nickel Fluosilicate, by R. P. PENROSE and K. W. H. STEVENS.

ABSTRACT. The paramagnetic resonance from the nickel ion in nickel fluosilicate has been investigated at a number of frequencies and temperatures. It has previously been suggested that the ion is in an electric field of trigonal symmetry. The results obtained are in good agreement with this hypothesis. It is also found that the magnitude of the splitting of the levels varies with temperature.

The Theory of Magnetic Resonance-Line Widths in Crystals, by M. H. L. PRYCE and K. W. H. STEVENS.

ABSTRACT. A theory is developed for calculating the mean displacements and mean square widths of resonant absorption lines in crystals for which the spin-lattice relaxation time is long compared with the spin-spin relaxation time. The theory is applied to a number of different cases; in particular, the effects of nuclear hyperfine structures and exchange forces in ionic resonances are discussed at length. It is shown that the theory holds, provided that the temperature is high compared with the Curie temperature.

On the Capture of Particles into Synchrotron Orbits, by T. R. KAISER.

ABSTRACT. A theory is developed for the capture of particles into synchrotron orbits when the radio-frequency accelerating voltage rises from zero to its maximum value in a finite time. It is shown that if this time of rise is sufficiently long, all particles originally occupying a band of width $1/\sqrt{2}$ times the maximum width of the final stable region for synchrotron phase oscillations will be captured, irrespective of the initial phase at which they enter the accelerating gap. This conclusion would seem to be in agreement with the observed characteristics of existing electron synchrotrons.

Experiments on Electron Capture and Phase Stability in a 14 mev. Synchrotron, by T. R. KAISER and J. L. TUCK.

ABSTRACT. Experiments were made by interrupting the radio-frequency accelerating voltage for intervals during the acceleration in a synchrotron, and observing the effects on the output. The process of capture is found to be more efficient than would be expected for an instantaneously developed radio-frequency accelerating voltage. The synchrotron acceleration may be interrupted for a short period without losing more particles than would be accounted for by contraction of the beam radius during the interruption, so that for sufficiently short periods of interruption the loss is negligible. This demonstrates that the phase at which electrons enter the resonator during capture is of no consequence under the conditions prevailing. The results are in quantitative agreement with a theory put forward by Kaiser which takes into account the finite rate of rise of resonator voltage, and with the general theory of synchrotron stability.

THE PROCEEDINGS OF THE PHYSICAL SOCIETY

Section B

VOL. 63, PART 2

1 February 1950

No. 362 B

The Properties of Ferromagnetic Compounds at Centimetre Wavelengths

By J. B. BIRKS

Department of Natural Philosophy, University of Glasgow

MS. received 29th June 1949

ABSTRACT. The magnetic and dielectric properties of γ -ferric oxide, magnetite, Mn-Zn ferrite and Ni-Zn ferrite have been measured at wavelengths from 60 cm. to 1.23 cm. The pronounced magnetic dispersion and absorption which occur are attributed to the decay of the rotational magnetization, due to Larmor spin resonance in the internal anisotropy field H_N . The theoretical relation $H_N = 2K/M$, where K is the anisotropy coefficient, and M the saturation intensity, agrees satisfactorily with the observations, and $|K|$ has been derived for each material. An inverse relationship between H_N and the initial susceptibility is shown to account quantitatively for the magnetic resonance observed in high-permeability materials at much lower frequencies.

§ 1. INTRODUCTION

AN investigation has been made of the magnetic and dielectric properties of four ferromagnetic compounds at wavelengths from 60 cm. to 1.23 cm., using the waveguide dual-impedance method, previously described (Birks 1948 a). The complex permeability and permittivity of the materials are obtained from observations of the input impedances of thin specimens, fitted into coaxial or rectangular waveguides, when terminated in a short- and an open-circuit. The ferromagnetic compounds are of interest because, due to their low conductivity, the high-frequency field is able to penetrate much more deeply than in a metal. It is thus possible to observe magnetic dispersion and absorption effects due to internal phenomena, which are obscured in metals by the intense skin-effect.

Almost all the previous experimental work on high-frequency permeability has been concerned with metals (Allanson 1945) and several attempts have been made to interpret the observed dispersion in terms of various internal relaxation and resonance mechanisms (Arkadiew 1945), analogous to those occurring in polar dielectric materials (Debye 1929) and in paramagnetic salts (Gorter 1947). The major magnetic dispersion in ferromagnetic metals, which occurs at frequencies above 100 Mc/s., is not, however, explicable in such terms, as the skin-effect restricts the penetration of the oscillatory field to a thin surface layer of less than a single domain thickness. Kittel (1946) has proposed an alternative magnetic dispersion mechanism for metals, based on this incomplete penetration of surface domains. Measurements on the microwave properties of carbonyl iron powder give results in agreement with Kittel's theory (Birks 1948 b).

Since the microwave magnetization of metals is only a surface effect, it appears that if high-frequency measurements are to be used to study the internal dynamics of ferromagnetic materials, they must be made on materials of conductivity sufficiently low for the skin-depth to be large compared with the domain size. The ferromagnetic compounds— γ -ferric oxide, magnetite and the ferrites—satisfy this criterion. The magnetic dispersion and absorption observed in these materials can hence only arise from internal restraints on the magnetization (Birks 1948 a).

§ 2. EXPERIMENTAL RESULTS

(i) *Typical γ -ferric Oxide Mixture*

The ferromagnetic compounds, in the form of finely-divided powders, were mixed with paraffin wax to form solid samples for measurement. A typical series of observations on a single mixture, containing 32.6% by volume of γ -ferric oxide in wax, will be described initially to illustrate the general nature of the results. The complex permeability $\mu = \mu' - i\mu''$ and the complex permittivity $\epsilon = \epsilon' - i\epsilon''$ of the mixture were measured at 9 different wavelengths: 58.5, 39.2, 29.8, 22.4, 15.3, 8.93, 5.97, 3.09 and 1.23 cm. Measurements were made at each wavelength on three or more specimens, moulded from the same material. The mean values of the components of μ and ϵ are plotted in Figures 1 and 2. The initial static permeability μ_s of the mixture, measured by a ballistic method, was 3.2.

Although the magnetic properties are of primary interest, the simultaneous derivation of ϵ from the same set of observational data provides a check on the validity of the measurements. The gradual decrease of permittivity observed is typical of the dielectric dispersion of solid inorganic compounds. Unusual features in the magnetic dispersion cannot, therefore, be attributed to experimental error. This point is emphasized, because certain abnormal magnetic dispersion phenomena reported in metals (Kartschagin 1922, Gans and Loyarte 1921) were later shown to arise from errors of observation (Wait 1927).

The magnetic dispersion and absorption (Figure 1) is pronounced, and displays several interesting features. On the long-wavelength side of the dispersion region ($\lambda \sim 150$ cm.), the limiting value of μ' , $\mu_0 = 3.05$ is within 5% of μ_s , the static permeability, and hence no appreciable dispersion can occur in the interval from $\lambda = \infty$ to $\lambda \sim 150$ cm. At this latter wavelength μ' begins to decrease, becoming unity at $\lambda \sim 7.5$ cm., while $\tan \delta_\mu (= \mu''/\mu')$ rises to a maximum at about the same wavelength. At shorter wavelengths, $\tan \delta_\mu$ decreases again rapidly, and μ' falls below unity, corresponding to a negative susceptibility. At $\lambda = 1.23$ cm. the magnetic loss is negligible, while μ' is approximately unity, which is the ultimate infra-red value for ferromagnetic materials (Hagen and Rubens 1903). Thus, practically all the magnetic dispersion and absorption occurs in the wavelength interval 150 cm. to 1 cm.

Further measurements were made to determine the effect of a static magnetic field on the observed dispersion. Thin discs (thickness/diameter ≈ 0.1) of the materials were fitted normally in the coaxial line impedance apparatus, and the static magnetic field H was applied parallel to the faces of the discs. Specimens of similar dimensional ratios were used to avoid differences in demagnetization coefficients. The effective values of μ as a function of H , up to 2,500 gauss, were measured at wavelengths from 58.5 cm. to 5.97 cm. The permittivity ϵ was unaffected by H . The static magnetization of the mixture was also measured by a ballistic method, and the incremental static permeability $\mu_s (= dB/dH)$ and its reversible component μ_{sr} were derived as functions of the applied field H .

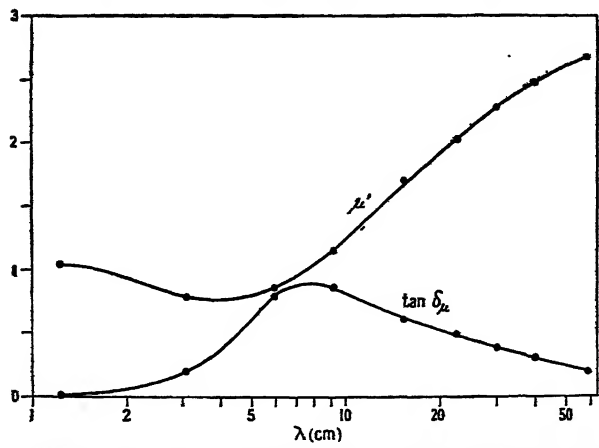


Figure 1. Magnetic dispersion and absorption of 32.6% γ -ferric oxide mixture.

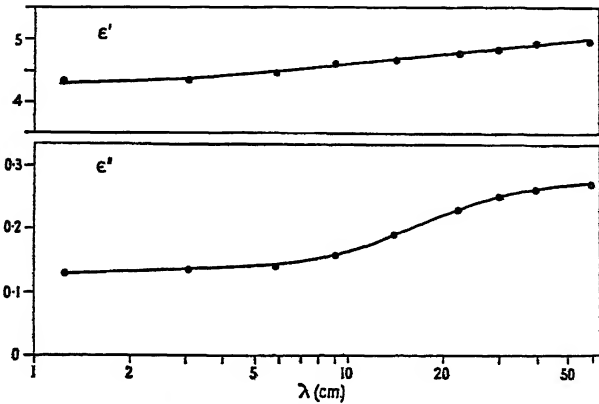


Figure 2. Dielectric dispersion and absorption of 32.6% γ -ferric oxide mixture.

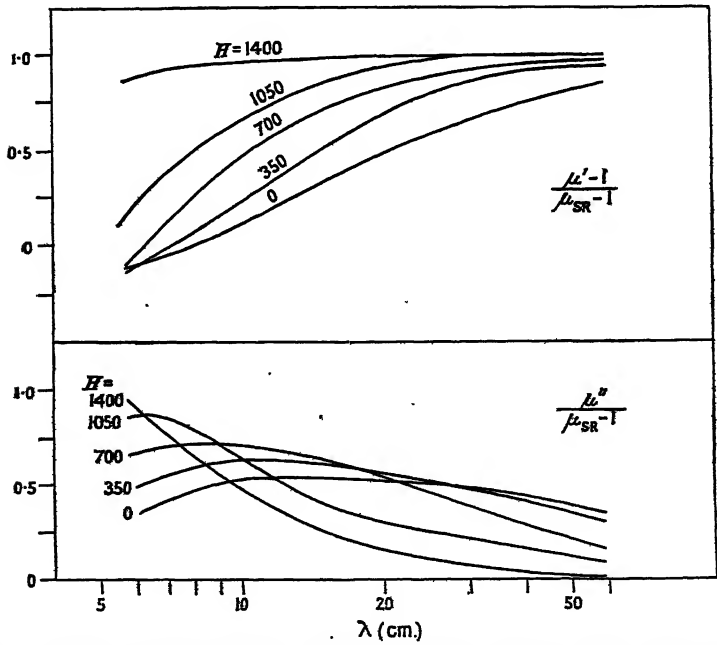


Figure 3. Effect of applied field H on magnetic dispersion (32.6% γ -ferric oxide mixture).

The results of the high-frequency and static measurements, adjusted to allow for the demagnetization coefficients of the coaxial specimens, have been published previously (Birks 1947). The derived magnetic dispersion and absorption curves, corresponding to different values of H , are plotted in Figure 3. For comparison, the ordinates are expressed relative to the corresponding values of $\mu_{SR} - 1$. It will be observed that the magnetic dispersion retains a similar form, but that it shifts to shorter wavelengths, and becomes steeper, as the applied field H is increased.

(ii) Magnetic Dispersion of Ferromagnetic Compounds

Magnetic and dielectric dispersion and absorption curves, similar to Figures 1 and 2, have been obtained for mixtures containing other concentrations of γ -ferric oxide. It has been found (Birks 1948 a) that, at each wavelength of measurement, μ and ϵ of the mixtures vary with v , the proportion by volume of the ferromagnetic compound, according to the relations

$$\log |\mu| = v \log |\mu_a|, \quad \dots\dots (1)$$

$$\tan \delta_\mu = v \tan \delta_{\mu_a}, \quad \dots\dots (2)$$

$$\log |\epsilon| = v \log |\epsilon_a| + (1-v) \log |\epsilon_w|, \quad \dots\dots (3)$$

$$\tan \delta_\epsilon = v \tan \delta_{\epsilon_a}, \quad \dots\dots (4)$$

where suffix a refers to the (extrapolated) properties of the ferromagnetic solid, and suffix w to the properties of the wax ($|\epsilon_w| = 2.28$, $|\mu_w| = 1.00$, $\delta_{\mu_w} = \delta_{\epsilon_w} = 0$.) These relations have been used to extrapolate the observations on the mixtures to zero dilution, and thus obtain the properties of the ferromagnetic solid. The derived magnetic dispersion and absorption curves of γ -ferric oxide are plotted in Figure 4.

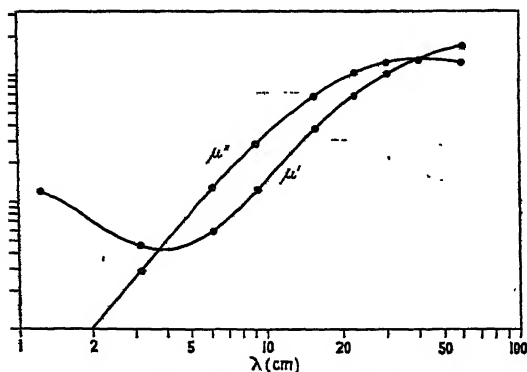


Figure 4. Magnetic spectrum of γ -ferric oxide.

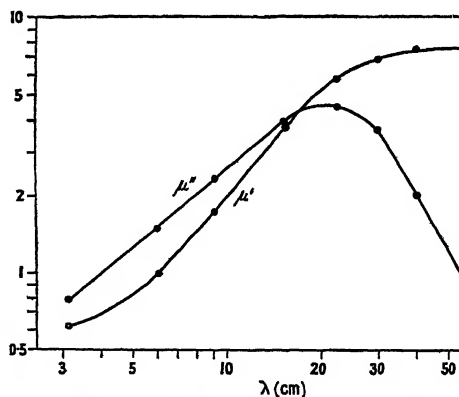


Figure 5. Magnetic spectrum of magnetite

Observations have been made on three similar series of mixtures containing various concentrations of magnetite, nickel-zinc ferrite and manganese-zinc ferrite respectively. The mixture laws (1)–(4) were again found to be valid. The derived magnetic spectra of these three materials are plotted in Figures 5, 6 and 7. The same type of broad magnetic resonance dispersion occurs in each compound.

The values of μ_0 , the limiting permeability at frequencies below the dispersion region, and of μ_s , the initial static permeability, have been obtained from the dispersion curves, and from static measurements on mixtures respectively, and they are listed in Table 1.

§ 3. THEORY OF MAGNETIC DISPERSION

(i) Nature of Magnetization Process

The domain theory of ferromagnetism (Becker and Döring 1939) differentiates four elementary processes which occur in the static magnetization of a demagnetized material.

- I. Reversible displacements of the Bloch walls separating adjacent domains.
- II. Irreversible displacements of the domain walls.
- III. Irreversible rotations of the spins within a domain from one direction of easy magnetization to another.
- IV. Reversible rotations of the domain spins towards the direction of the applied field.

The reversible translational magnetization I takes place in weak fields only, and is generally considered responsible for the initial permeability. The irreversible processes II and III, responsible for the Barkhausen effect, occur in

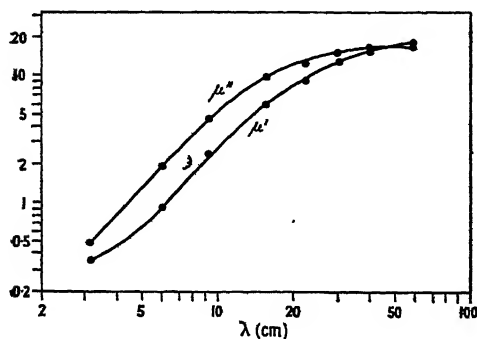


Figure 6. Magnetic spectrum of nickel-zinc ferrite.

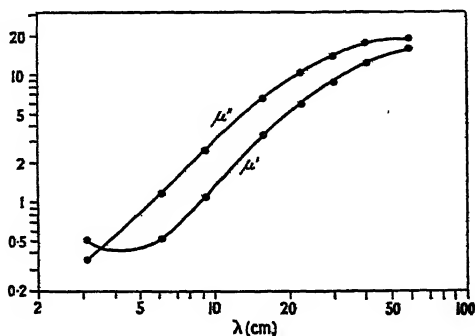


Figure 7. Magnetic spectrum of manganese-zinc ferrite.

Table 1

Material	μ_s	μ_0	λ_N (cm.)	ν_N (Mc/s.)	H_N (gauss)	M	$ K $ (ergs/cm ³)
Magnetite	15	8	18	1670	590	450	1.3×10^5
γ -ferric oxide	34	30	40	750	270	350	4.7×10^4
Ni-Zn ferrite	40	34	50	600	210	300	3.2×10^4
Mn-Zn ferrite	48	44	70	430	150	200	1.5×10^4

medium fields over the steep portion of the B - H curve. The reversible rotational magnetization IV can occur at all field strengths, and is the only process operative in high fields.

Becker (1938) has developed a theory of magnetic dispersion for the translational and irreversible magnetization components. The dispersion is attributed to the damping action of eddy current fields, induced in the vicinity of domains by the extension of their boundaries or by irreversible rotations. The disappearance of the irreversible magnetization components at relatively low frequencies, predicted by Becker, has been confirmed experimentally (Birks 1947). Hence, only reversible processes are operative in the microwave region. For the relaxation

of the reversible translational magnetization, Becker obtains a Debye-type equation of the form

$$\frac{\mu-1}{\mu_s-1} = \frac{1}{1-iv/\nu'} \quad \dots\dots (5)$$

for the permeability μ in weak oscillatory fields of frequency ν , where ν' is the relaxation frequency. The form of (5) is identical with the dispersion relation obtained by Arkadiew (1913) from his semi-empirical theory of magnetic viscosity. The observed dispersion and absorption curves differ appreciably from (5). The (μ', λ) curves are steeper, the (μ'', λ) curves are sharper, and $\mu' - 1$ becomes negative, while (5) gives positive values only. The permeability does not, therefore, decay by a relaxation process, as would be expected if it were due to translational magnetization.

The observations on the effect of an applied field (Figure 3) show that a similar type of dispersion occurs when high static fields are present. Process I, which is restricted to weak fields, cannot be responsible, and hence the magnetic dispersion must be accounted for in terms of the one alternative process, reversible rotational magnetization, which is operative at all field strengths.

Comparison of μ_0 and μ_s (Table 1) shows that, with the exception of magnetite, practically all the dispersion occurs in the microwave region. Since this dispersion is attributable to rotational magnetization, it follows that the translational magnetization in these materials, even in static fields, is relatively small. This is probably due to their low conductivity, which will inhibit the spread of eddy-currents associated with domain-boundary displacements. Magnetite, in which the initial rotational and translational components are of equal magnitudes ($\mu_s \simeq 2\mu_0$), has a higher conductivity than the other three compounds. The magnetic dispersion observed in magnetite by Zimowski (1937) at 5–20 Mc/s. probably corresponds to the decay of the translational magnetization.

(ii) *Natural Ferromagnetic Resonance*

The rotation of the spins within a domain towards the direction of the applied field is opposed by the crystalline anisotropy forces, which tend to keep the spins aligned along an easy magnetization axis, On . Landau and Lifshitz (1935), in considering a theoretical domain model of a ferromagnetic crystal, have represented these anisotropy forces by an equivalent internal magnetic field H_n , acting along On . This natural anisotropy field, which is of the order of 1,000 gauss, arises from spin-orbit interaction, and is distinct from the much larger Weiss molecular field, arising from the spin exchange forces. Landau and Lifshitz have considered the effect of a weak oscillatory magnetic field applied transverse to On , so that spin rotation is the only magnetization process possible. They find that magnetic dispersion takes place as a resonance process, resonance occurring at the Larmor precession frequency ν_n of the electron spins in the internal field, given by

$$h\nu_n = g\mu_B H_n, \quad \dots\dots (6)$$

where g is the Landé factor ($=2$ for electron spins), and μ_B is the Bohr magneton.

The differences, noted above, between the observed dispersion and the relaxation equation (5), and, in particular, the negative values of $\mu' - 1$, are characteristic of a resonance effect. These observations provide direct experimental evidence of the existence of an internal field. Any magnetic resonance

process occurring in the microwave region must be associated with a field of the order of 1,000 gauss, since for electron spins (6) becomes numerically

$$H_n \lambda_n = 10.7 \times 10^3 \text{ gauss cm.}, \quad \dots\dots(7)$$

where $\lambda_n (=c/\nu_n)$ is the resonant wavelength. Since magnetic resonance is observed when no such field is applied externally, an effective field H_n must exist within the ferromagnetic material. When an external field H is applied to the material, it will increase the total field acting on the spins, and the resonance will shift to higher frequencies, as observed (Figure 3). When $H > H_n$, the resonant frequency will be determined primarily by the external field, and induced ferromagnetic resonance will occur (Birks 1948 c). Thus, qualitatively, the observations are consistent with Landau and Lifshitz's hypothesis of the existence of an internal field. The identification of this field with that arising from anisotropy will be discussed later.

(iii) *Natural Resonance in Polycrystalline Materials*

The conditions within each domain or region of a polycrystalline material, in which a unidirectional internal field H_n exists, will be similar to those in a paramagnetic or saturated ferromagnetic material in which the spins are aligned by an external polarizing field H_n . In this latter case, when an oscillatory field is applied transverse to H_n , induced magnetic resonance occurs at the Larmor frequency ν_n of the spins in the external field (Zavoisky 1946). This effect is thus closely related to the natural magnetic resonance, due to the internal field. The resonance dispersion equation, derived by Frenkel (1945),

$$\frac{\mu - 1}{\mu_0 - 1} = \frac{\nu_n^2}{\nu_n^2 - \nu^2 + 2i\nu\nu'}, \quad \dots\dots(8)$$

where ν' is the damping frequency, μ_0 the initial rotational permeability, and μ the permeability at frequency ν , has been verified experimentally for a saturated ferromagnetic by Yager and Bozorth (1947). The experiments on induced ferromagnetic resonance show that $\nu' \ll \nu_n$, and, hence, from (8), the maximum absorption μ'' occurs when $\nu = \nu_n$.

In an unmagnetized polycrystalline material the internal field H_n will not be uniform, but will vary in direction, due to local changes in domain and crystal structure, and in magnitude, due to magnetic and thermal interactions. These will lead to a broad distribution of values of H_n about a mean H_N , corresponding to the undisturbed, uniaxial field. The natural polycrystalline magnetic dispersion thus consists of a broad spectrum of relatively sharp resonances ($\nu' \ll \nu_n$) occurring locally, and distributed about a mean resonant frequency ν_N . The over-simplification of applying (8) directly to the dispersion (Birks 1948 c)* is invalid, since it involves postulating values of $\nu' > \nu_n$ to account for the breadth of the resonance. In the absence of a suitable theory for treating the magnetic interactions, no attempt has yet been made to extend (8) to the polycrystalline case, though a tentative approach to the problem has been made recently by

* The larger values of H_n obtained previously, on the assumption, from (8), that the resonant wavelength is that at which $\mu' = 1$, were in error by a factor of ~ 10 . The corrected anisotropy fields are insufficient to account for more than a small fraction of the g -value anomaly ($g > 2.0$) observed in induced ferromagnetic resonance (Yager and Bozorth 1947). This anomaly is a distinct effect, and is not, as previously suggested (Birks 1948 c), due merely to anisotropy.

Crouch (1949). The mean resonant wavelength λ_N can be derived directly from the observed peak of magnetic absorption. The values of λ_N , ν_N and H_N obtained are given in Table 1.

The observations on the effect of an applied field (Figure 3) agree qualitatively with the theory proposed. Apart from the shift towards higher frequencies due to the increase of the effective field, there is a reduction in the breadth of the resonance, corresponding to the decrease in the magnetic disorder in the material as the domain moments are turned irreversibly from random orientations towards the direction of the applied field.

§ 4. THE INTERNAL FIELD

(i) Anisotropy Energy and the Internal Field

The anisotropic properties of a ferromagnetic material are normally described in terms of an anisotropy energy coefficient K . For a cubic crystal, magnetized to saturation, the component of the magnetic energy due to anisotropy may be written

$$f = K(\alpha_1^2\alpha_2^2 + \alpha_2^2\alpha_3^2 + \alpha_3^2\alpha_1^2), \quad \dots\dots(9)$$

where α_1 , α_2 , α_3 are the direction cosines of the magnetization relative to the (100) axes.

The internal anisotropy field H_N is related to K and the saturation intensity M . Let us consider a single domain of a cubic crystal, lying in the [100] plane Oyz , whose easy direction of magnetization lies along a (100) axis, Oz (Figure 8). The

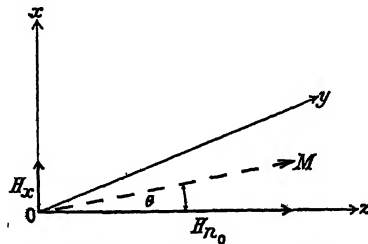


Figure 8. Rotational magnetization process in a single domain.

Note. H_{n_0} in abscissa should read H_N .

internal field H_N acts along Oz , and in the unmagnetized state M is in the same direction. If a small transverse field H_x is applied along Ox , M will rotate through an angle θ in the plane Oxz . The internal field H_N exerts a restoring torque on M , given by

$$T = MH_N \sin \theta. \quad \dots\dots(10)$$

Alternatively, if the process is described in terms of f , in the plane Oxz ,

$$f = K \cos^2 \theta \sin^2 \theta \quad \dots\dots(11)$$

and the anisotropy energy exerts a restoring torque

$$T = \frac{\partial f}{\partial \theta} = K \sin 2\theta \cos 2\theta. \quad \dots\dots(12)$$

Equating (10) and (12), for θ small, we obtain

$$H_N = \frac{2K}{M}. \quad \dots\dots(13)$$

For a crystal in which the (111) axes are preferred, a similar analysis gives

$$H_N = -\frac{2K}{M}. \quad \dots\dots(14)$$

The value of K is normally derived experimentally from the differences between the energies of magnetization of a single crystal along its different axes. Quittner (1909) has measured the magnetization curves of a magnetite crystal, and a value of $K = -1.124 \times 10^5$ ergs/cm³ has been derived from his data. Magnetite, like the other compounds studied, has a spinel-type cubic structure. Its preferred direction of magnetization is along the cube diagonal. The saturation intensity of artificial magnetite at room temperature, $M=450$ (Weiss and Forrer 1929) and hence from (14) the theoretical value of the anisotropy field $H_N=500$. This agrees satisfactorily with that for the internal field, $H_N=590$, obtained from the natural magnetic absorption peak.* It is concluded therefore that the internal field responsible for the magnetic resonance does arise from the ferromagnetic anisotropy.

No data are available on the magnetic properties of single crystals of the other compounds studied. The magnitude of the anisotropy constant $|K|$ can be estimated from the experimental values of H_N and M , using (13) or (14). The sign of K will depend on the preferred directions of magnetization. The values of $|K|$ obtained are listed in Table 1. The magnitude of K is one of the principal factors influencing the initial permeability, a small $|K|$ being associated with a large μ_s . It will be observed from Table 1 that the two quantities are found to be in inverse sequence.

(ii) *The Initial Permeability and the Internal Field*

Comparison of the initial permeabilities of the four compounds with those from other sources (Welo and Baudisch 1925, Kittel 1946, Snoek 1947) indicates that, with the exception of the manganese-zinc ferrite, the specimens are representative. The properties of the Mn-Zn ferrite specimen, however, differ appreciably from the optimum mixtures obtained by Snoek (1947, 1948), which have $\mu_s \approx \mu_0 \approx 1,000$. In such materials $|K|$ will be much smaller, and, consequently, natural magnetic resonance should occur at much lower frequencies. Snoek (1948) has observed the initial part of a sharp rise in the magnetic absorption of optimum Mn-Zn ferrites at frequencies below 1 Mc/s. This is interpreted by Snoek as the fringe of a magnetic resonance region, but his published observations are confined to frequencies well below the natural resonance frequency. From the limited data it is estimated that this is ≈ 20 Mc/s., which is equivalent, on the theory given above, to $H_N \approx 7$ gauss and $|K| \approx 700$ ergs/cm³.

A simple model, similar to that of Figure 8, has been used by Snoek (1948) to relate the initial rotational susceptibility X_0 and the internal field H_N . Neglecting all interactions, and assuming that H_N is constant throughout the polycrystalline material, it is found that

$$H_N = \frac{2M}{3X_0} = \frac{8\pi M}{3(\mu_0 - 1)}, \quad \dots\dots(15)$$

* The residual discrepancy between theory and experiment is reduced, if a value of g greater than 2.0, as indicated by induced ferromagnetic resonance experiments (Yager and Bozorth 1947, Birks 1948 c) is adopted; e.g. $g=2.2$, $H_N=536$; $g=2.36$, $H_N=500$.

so that for a given value of M , H_N should be inversely proportional to $\mu_0 - 1$. The parameters derived for the ferrite, $\mu_0 \approx 1,000$, can be compared with those for the specimen, $\mu_0 \approx 44$. The values of H_N are approximately in the ratio 1 to 22, while those of $\mu_0 - 1$ are in the ratio 23 to 1. Thus the relative magnitudes of the internal fields and the resonant frequencies are simply accounted for by the difference in the initial permeabilities of the two specimens.

In Table 2 the theoretical values of H_N from M , μ_0 and (15) are compared

Table 2

Material	Expt.	H_N		μ_0
		Theory (15)	$\frac{H_N(\text{exp})}{H_N(\text{theor})}$	
Magnetite	590	540	1.1	8
γ -ferric oxide	270	100	2.7	30
Ni-Zn ferrite	210	74	2.9	34
Mn-Zn ferrite	150	42	3.6	44

with the experimental results. There is close agreement in the case of magnetite, but for the other materials there is a discrepancy by a factor of the order of 3. This order of agreement is satisfactory, however, considering the extreme simplicity of the theoretical model and the neglect of all interactions in the derivation of (15).

ACKNOWLEDGMENTS

The author wishes to acknowledge the tenure of an I.C.I. Research Fellowship during the initial stages of this work, and also to thank Professor P. I. Dee for the use of the facilities of his laboratory.

REFERENCES

- ALLANSON, J. T., 1945, *J. Instn. Elect. Engrs.*, **92**, III, 247.
 ARKADIEV, W., 1913, *Phys. Z.*, **14**, 928 ; 1945, *J. Phys. U.S.S.R.*, **9**, 373.
 BECKER, R., 1938, *Phys. Z.*, **39**, 856.
 BECKER, R., and DÖRING, W., 1939, *Ferromagnetismus* (Berlin : Springer).
 BIRKS, J. B., 1947, *Nature, Lond.*, **160**, 535 ; 1948 a, *Proc. Phys. Soc.*, **60**, 282 ; 1948 b, *Phys. Rev.*, **74**, 843 ; 1948 c, *Ibid.*, **74**, 988.
 CROUCH, G. E., 1949, *Phys. Rev.*, **75**, 525.
 DEBYE, P., 1929, *Polar Molecules* (New York : Chemical Catalog Co.).
 FRENKEL, J., 1945, *J. Phys. U.S.S.R.*, **9**, 229.
 GANS, R., and LOYARTE, R. G., 1921, *Ann. Phys., Lpz.*, **64**, 209.
 GORTER, C. J., 1947, *Paramagnetic Relaxation* (Amsterdam : Elsevier).
 HAGEN, E., and RUBENS, H., 1903, *Ann. Phys., Lpz.*, **11**, 873.
 KARTSCHAGIN, K., 1922, *Ann. Phys., Lpz.*, **67**, 325.
 KITTEL, C., 1946, *Phys. Rev.*, **70**, 281.
 LANDAU, L., and LIFSHITZ, E., 1935, *Phys. Z. Sowjet*, **8**, 153.
 QUITTNER, V., 1909, *Ann. Phys., Lpz.*, **30**, 289.
 SNOEK, J. L., 1947, *New Developments in Ferromagnetic Materials* (Amsterdam : Elsevier) ; 1948, *Physica*, **14**, 207.
 WAIT, G. R., 1927, *Phys. Rev.*, **29**, 566.
 WEISS, P., and FORRER, R., 1929, *Ann. Phys., Paris*, **12**, 279.
 WILCO, L. A., and BAUDISCH, O., 1925, *Phil. Mag.*, **50**, 399.
 YAGER, W. A., and BOZORTH, R. M., 1947, *Phys. Rev.*, **72**, 80.
 ZAVORSKY, E., 1946, *J. Phys. U.S.S.R.*, **10**, 197.
 ZAVORSKY, E., 1947, *J. Phys. U.S.S.R.*, **11**, 6.

The Field in an Electron-Optical Immersion Objective

By L. JACOB

I.C.I. Fellow, The University, Manchester *

*Communicated by J. Rotblat ; MS. received 4th August 1948,
in amended form 7th July 1949*

ABSTRACT. The electrolytic tank method was developed to record rapidly the potential variations which occur in the field of the immersion objective in the absence of space charge. The axial distribution over the region explored was found to conform with a law of the form $V = Ae^{kz}$. Here A is a voltage scale factor while k is a geometrical scale factor. The experimental results appear to show that for constant black-out voltage the gradient at the cathode remains constant over part of the range explored, independent of geometry. The ratio of accelerator to modulator voltage for zero gradient at the cathode varies inversely as the square of the hole diameter. A determination of the emitting area of the cathode was also made from the field plots.

§ 1. INTRODUCTION

THE electron optical immersion objective derives its name from the immersion objective of 'light' optics in that the electron-emitting area is immersed in, and forms part of, the lens field. In a wide sense it forms an integral part of every electron-emitting device in which electrons are drawn away to one or more electrodes by an electric field. Some of its focal properties, including aberration effects, have been described by Johansson (1933), and a study of its action as a gun for cathode-ray tubes can be found in the book by Maloff and Epstein (1938). Recently, very useful empirical design curves have been published by Moss (1945) which enable beam characteristics to be controlled.

The fundamental problems involved are those associated with the behaviour of electrostatic fields towards electron beams entering them from a cathode source. When these fields are suitably arranged, they have a focusing action on the electron beam; actually the field can act as a lens or a mirror. In the former case they produce a real 'cross-over' or minimum section of the beam within the lens field, and a cathode image further away. Moreover, they have a focal length which is necessarily small, and related to the distances of object and image by a slight modification of the usual lens formula.

The present study of these properties is mainly concerned with establishing the form of the field under various conditions of modulation. The resulting gradients at the cathode surface in a field which is space-charge free are worked out for the different cases. The effect of variations in geometry on the distribution are described since the information yielded has a direct bearing on the design of electron guns.

§ 2. METHOD OF EXPERIMENT

The field-distribution plots along and off the axis were obtained using an electrolytic tank in which distilled water was the conducting medium. Half-section ten-scale models of cathode, modulator and accelerator, which composed the objective, were silver-plated and suspended from two parallel ebonite

* Now at University of Liverpool.

rods attached to the removable tank-cover of brass. The probe was a silver wire, 0.2 mm. in diameter, fixed in a small Perspex chuck, which was rigidly connected to the movable screw-thread S. Figures 1(a) and 1(b) show the tank and electrode

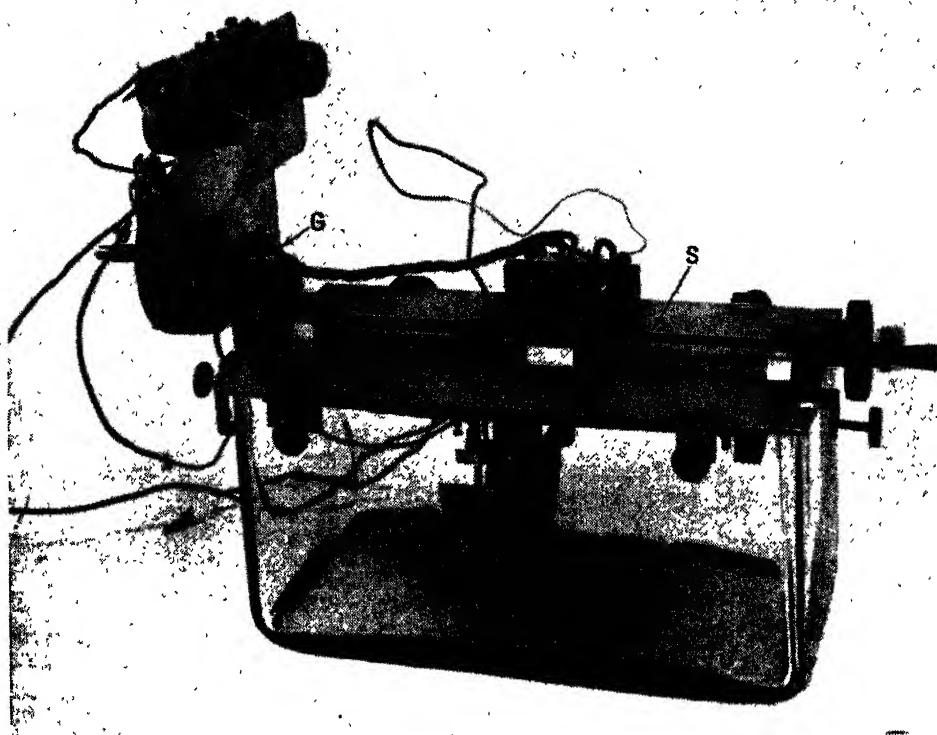


Figure 1 (a). Half-section ten-scale model in tank.

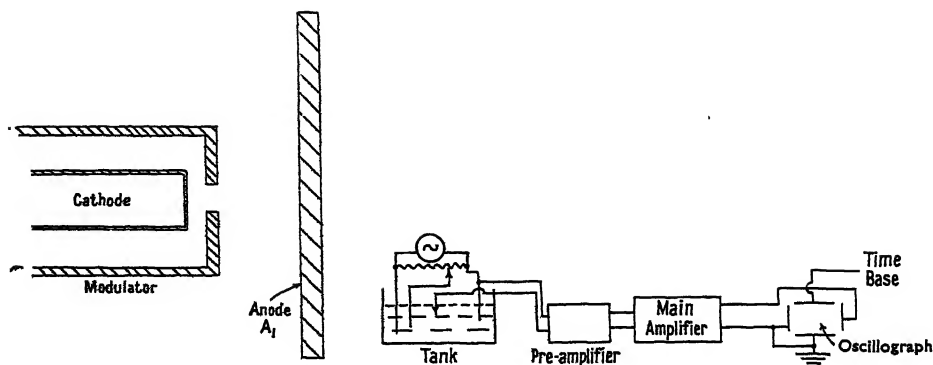


Figure 1 (b). Electron optical system.

Figure 2.

arrangement respectively. The set-up was similar to that described in an earlier paper (Jacob 1938) and the same precautions had to be strictly observed. These early tank plots were obtained by using a null method to trace out the equipotentials which were then recorded by an electromagnetic pencil placed vertically above

the probe. While the results obtained were accurate to 3%, the method was tedious in practice, requiring several hours for the complete plot. In the present case it was decided to work mainly on the axial distribution and to make the recording semi-automatic. This is achieved by picking up the potential difference between the probe and one electrode, applying it first to a pre-amplifier, and then to the main amplifier, whose output is fed to one pair of plates of an auxiliary oscillograph (Figure 2). The other pair of plates has a D.C. linear time-base potential, synchronized with the movement of the probe, through a system of gears *G* linked to the brass spindle *S*. In this way the potential distribution is seen as a fluorescent trace on the face of the cathode-ray tube.

Where a large number of traces are required—for example, when investigating the effect of variation of parameters such as cathode/modulator (C/M) spacing, anode/modulator (A/M) spacing, etc.—the method has proved expeditious, usually taking one minute per distribution. Adjustment to the electrodes was facilitated by removal of the brass cover top from which they were suspended.

One special feature of the present method was the ability to amplify any small portion of the path of the probe independently of the rest of the path. This was especially useful where it was desired to investigate the conditions obtaining close to the cathode when the slope of the field was small. The evaluation of the gradient then became practicable, since magnifications of ten or more could be used. Owing to the rectifying action of the pre-amplifier, the traces are, of course, recorded as D.C. on one side of the base-line only, and it is necessary in interpreting the plots to bear this in mind. The traces were photographed and enlarged for any subsequent calculations of gradients, emitting area, and so on.

§ 3. THE AXIAL POTENTIAL DISTRIBUTION

The chief region of interest in the axial plot lies between the cathode and some point up to or beyond the modulator for which the distribution of field is non-uniform (it was found in the preliminary work that between this point and the anode the potential rise was practically linear). It was hoped, by concentrating attention on this region, that a law might be found to enable the potential to be expressed as a function of distance from the cathode. Such a relation would simplify both calculations for the electron trajectory and those for establishing the cardinal points of the lens system. With this in view, families of curves were obtained for the axial distribution under the condition that the modulator was at cathode potential.

These are shown in Figure 3 for different C/M spacings, where the unit of distance is in terms of the modulator hole diameter, and the curve generally taken well beyond the position of the modulator. A plot of $\log V$ against distance yields a corresponding family of curves which are approximately straight lines, Figure 4, and consequently can be expressed in the form $V = Ae^{kx}$ over the limited region explored, where A and k are constants associated with the voltage and geometry respectively. The Table gives their values for the various C/M spacings for this type of system, where the modulator hole diameter is again taken as the unit of length.

C/M	0.05	0.10	0.20	0.30	0.50
<i>A</i>	0.056	0.066	0.069	0.058	0.018
<i>k</i>	1.61	1.30	0.83	0.74	0.78

In order to satisfy the boundary condition at the cathode ($V=0$), it is necessary to express the potential as a sinh function given by

$$V = A(e^{kx} - e^{-kx}) \quad \dots\dots(1)$$

and assume that the value of k remains unchanged as the cathode is approached. This means that a uniform field must exist in the immediate vicinity of the cathode; this soon merges into the simple exponential field derived from the field plots, while further along the axis towards the anode the distribution becomes linear again.

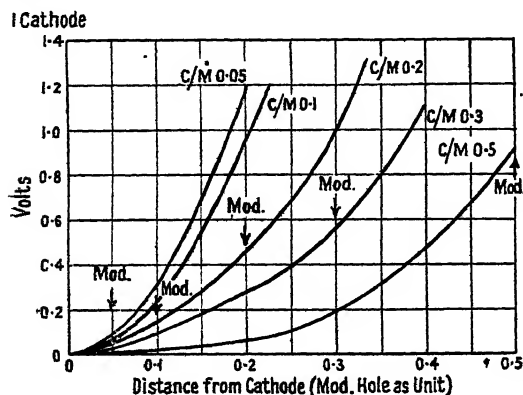


Figure 3.

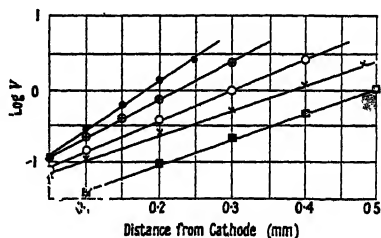


Figure 4.

- C/M 0.05 mm.: $V = 0.056 e^{1.61z}$
- ⊕ " 0.10 mm.: $V = 0.066 e^{1.3z}$
- " 0.20 mm.: $V = 0.069 e^{0.83z}$
- × " 0.30 mm.: $V = 0.058 e^{0.74z}$
- " 0.50 mm.: $V = 0.0182 e^{0.78z}$

Such a form of distribution along the axis can be derived by considering the solution of the Laplace equation for rotationally symmetric systems:

$$\frac{1}{r} \frac{\partial}{\partial r} \left(r \frac{\partial V}{\partial r} \right) + \frac{\partial^2 V}{\partial z^2} = 0. \quad \dots\dots(2)$$

When the potential V is put in the form $V = (\phi) \times (W)$, where ϕ is a function of r only and W is a function of z only, the solution for points on the axis takes the form

$$W = \Sigma \{ A_n \exp(k_n z) + B_n \exp(-k_n z) \}. \quad \dots\dots(3)$$

It thus might be possible, under certain conditions, to predict a form of exponential law as established empirically by the field plots.

In order to examine the early stages of the distribution in greater detail and to determine the effect of varying the modulator potential with respect to the cathode, use was made of the special amplifying facilities available. Three cases were investigated: (a) modulator at a positive potential; (b) modulator negative, up to that value which neutralizes the field from the anode—the cut-off or 'black-out' voltage V_B ; (c) modulator negative, beyond the 'black-out' point. These are shown in Figures 5(a), (b) and (c) respectively. At the 'black-out' point there is practically zero field at the centre of the cathode; the limiting cut-off current which is used in visual checking with sealed-off tubes must be made up of those electrons which, on account of their initial velocities, pass over the small negative barrier immediately in front of the cathode. Figure 5(c) shows that beyond the 'black-out' point there is a reversal of curvature on the actual traces. This arises from the rectifying action of the amplifier, so that the potential distribution is recorded in one

In the diagram, the positive direction has been taken downwards. To avoid confusion, only one of the curves, e.g. $V_M = -4.3$ v., is shown as it should appear. The negative minimum between the zero equipotential and the cathode indicates that the field has now taken the saddle-back form.

The exponential form of the region in the vicinity of the cathode can be deduced from Figure 6 for the different (negative) modulation voltages applied.

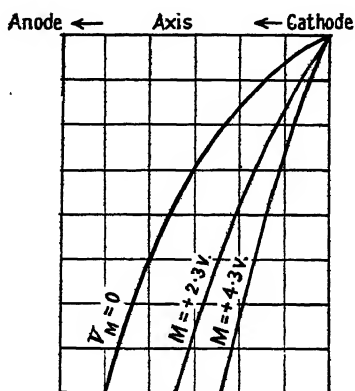


Figure 5 (a).

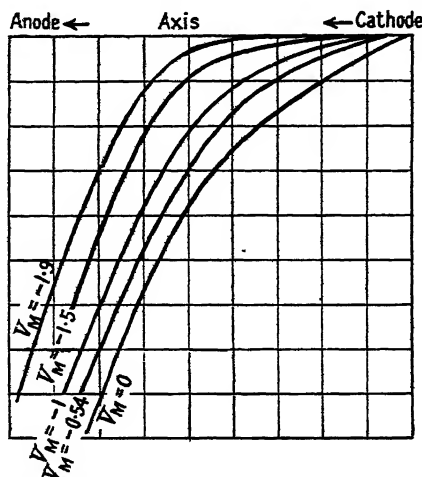


Figure 5 (b).

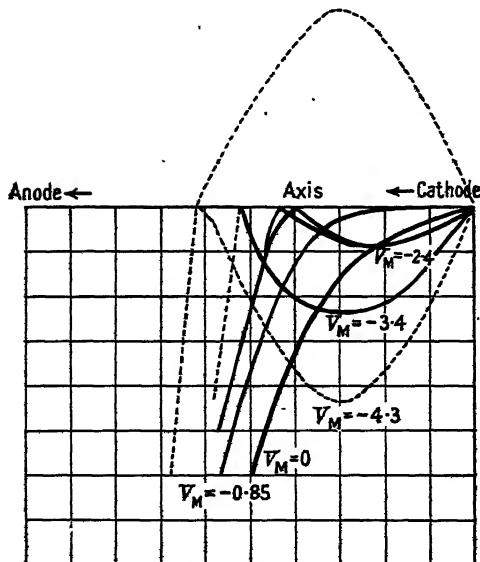


Figure 5 (c).

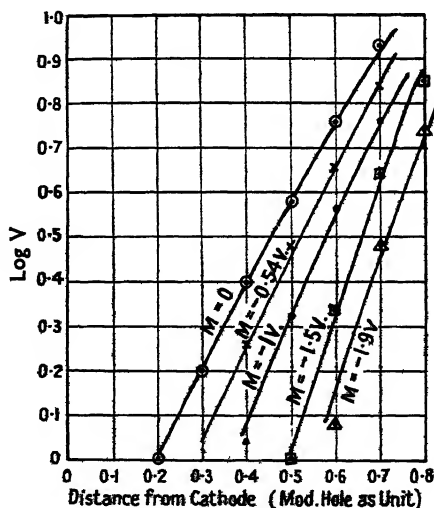


Figure 6.

Here again the curves are approximately straight lines with a slope increasing gradually as the modulator becomes more and more negative towards and past the 'black-out' point. The gradient at any point on the axis is thus proportional to the potential at that point, at least for static conditions. At the cathode surface the gradient is, of course, not zero, except in very special cases. Its value is $2kA$, obtained by differentiating equation (1) with respect to z . An analysis of

the curves of Figures 5 (a) and 5 (b) shows that the linear law holds only in cases of positive gradient all the way to the cathode. Where, as in Figure 5 (c), the field reverses in the neighbourhood of the modulator, the relationship becomes somewhat complicated.

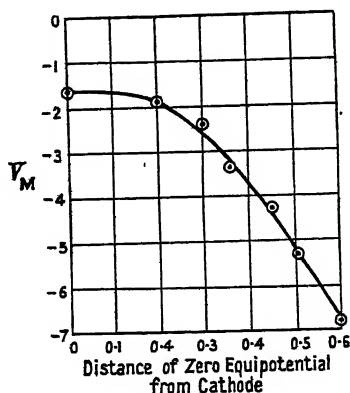


Figure 7.

The potential at any fixed point on the axis decreases with decrease in modulator potential; on the other hand, a point at a fixed potential moves away from the cathode under these conditions. This is illustrated in Figure 7, which shows the shift of the zero equipotential along the axis as the potential of the modulator is decreased; the effect is most marked for the early stages of modulation.

§ 4. GRADIENT AT CATHODE

Two important experimental relationships which affect the current-emitting properties of the cathode can be deduced from the field plots. These are (i) the gradient at the cathode, (ii) the point in the radial direction along the cathode surface at which the gradient normal to it vanishes; this defines the emitting area.

It was usual to work with the modulator at cathode potential for determinations of the slope. Let this be S volts/cm. The modulator was then biased back negatively to a value V_B in order to neutralize the effects of the anode field. By combining these values, the slope *in vacuo* for any value of the black-out voltage V_M is then $10S(V_M/V_B)$. Figure 8 (a) shows the average gradient at the cathode at $V_M=0$, calculated for a constant black-out voltage of 30 volts. This has been plotted against the C/M spacing for two different hole diameters. It will be seen that over a certain range the gradient at the centre of the cathode is constant at about 200 volts/cm. independent of the modulator hole diameter. For the larger spacings it tends to fall off rapidly for both modulators. The dotted line represents the gradient at the mid-point of the system. This increases approximately linearly with increase in C/M spacing for constant black-out voltage.

It is clear that the gradient will depend on the black-out voltage alone, provided that the anode voltage and geometry remain fixed in any one case.

The fact that kA is constant at the centre of the cathode for constant black-out voltage indicates the existence of a uniform field in this region. The 'cross-over' or minimum cross section of the beam is formed further away—in the exponential part of the field which extends to the other side of the modulating electrode, and which appears to govern the focusing properties of the lens.

The decrease in gradient at constant anode voltage is shown in Figure 8(b). The ratio of gradients is in proportion to the square of the modulator hole diameter. As expected, the thinner modulator gives the larger gradient, but there does not appear to be any simple relation involving the thickness of the modulator alone. An interesting curve is that shown in Figure 9, where the ratio of anode to modulator

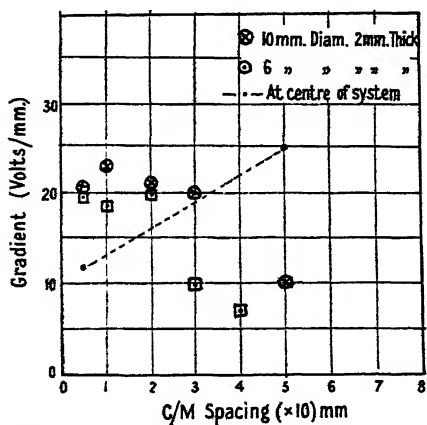


Figure 8 (a). Average gradient at cathode for $V_{BO}=30$ volts.

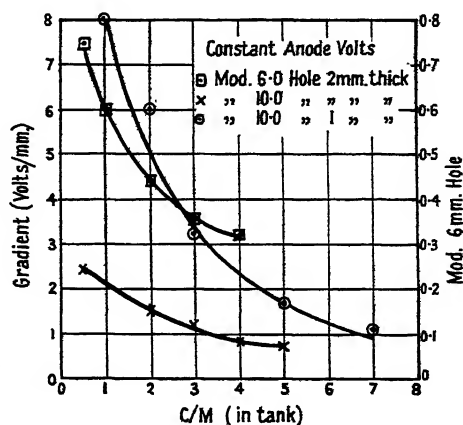


Figure 8 (b). Average gradient at cathode at $M=0$.

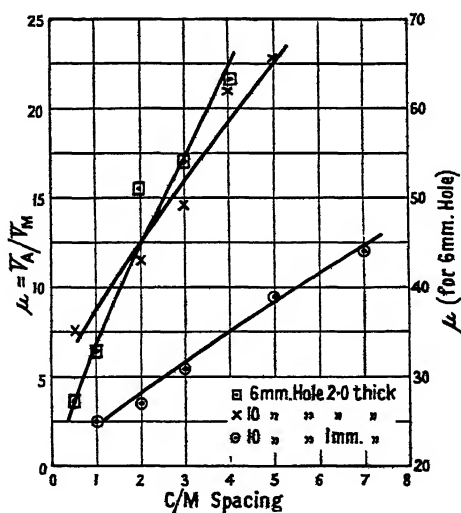


Figure 9.

voltage, V_A/V_M , for zero gradient at the centre of the cathode (i.e. for the black-out condition) is plotted against C/M spacing. The result is a linear relationship, as might have been predicted from orthodox theory. This ratio varies approximately inversely as the square of the hole diameter. From this we may infer that at constant anode voltage the gradient at the cathode surface is proportional to the modulator potential, and hence varies as the black-out voltage.

A useful curve for design purposes is that connecting the black-out voltage V_B with C/M spacing, since the current yield can be calculated from the value

of V_B . This is shown in Figure 10 (a) and (b) for the 10 mm. and 6 mm. modulator

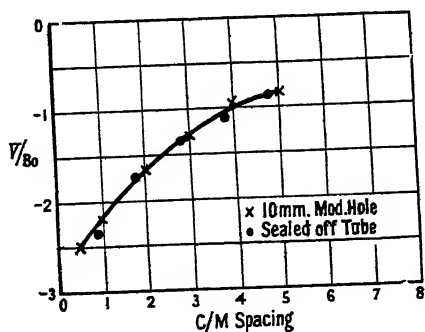


Figure 10 (a).

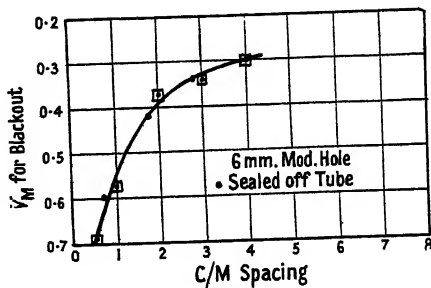


Figure 10 (b).

holes respectively. It will be noticed that the rate at which the black-out voltage falls is greatly increased at the short spacings; this necessitates special care in assembly to avoid a large spread in V_B . In order to check the accuracy of the static determination in the tank, a number of sealed-off gun units were made up to one-tenth the tank scale. These were pumped and processed in the normal way, and the black-out point was determined in each case by the disappearance of the fluorescent spot on the screen. The values were then reduced so as to make only one point correspond to a point on the curve; the remaining points are seen to lie fairly close to the tank curve.

In general, the conclusions reached from a study of the results on the 10 mm. aperture modulator could be applied to the 6 mm. aperture; for example, the exponential character of the law connecting potential and distance from the cathode was upheld in all cases of C/M variations, e.g. the shift of the zero equipotential along the axis with modulation.

§ 5. THE EMITTING AREA

It was possible to explore the gradient along the cathode surface at points off its centre by moving the probe along it. The point at which the gradient

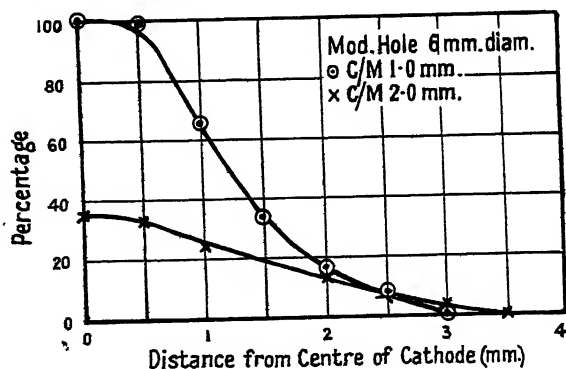


Figure 11. Average gradient as percentage of gradient at centre.

normal to the surface becomes zero defines the maximum emitting area of the cathode. This is shown in Figure 11 for the case when the modulator is at cathode potential for each of two separate C/M variations.

The ordinates when plotted as a percentage of the gradient at the centre of the cathode show a markedly increased rate of fall for the smaller spacing. The emitting area will be seen to correspond approximately with the projection of the modulator hole on to the cathode surface when, as in this case, the modulator electrode is tied to the cathode potential. For negative bias values the emitting area is proportional to the ratio of bias voltage to black-out voltage.

REFERENCES

- JACOB, L., 1938, *Phil. Mag.*, **26**, 570.
 JOHANNSON, H., 1933, *Ann. Phys., Lpz.*, **18**, 385.
 MALOFF, I. G., and EPSTEIN, D. W., 1938, *Electron Optics in Television* (New York and London: McGraw-Hill).
 MOSS, H., 1945, *J. Brit. Instn. Radio Engrs.*, **58**, 10,

Colour Perception in Parafoveal Vision

By M. GILBERT *

Imperial College of Science and Technology, London S.W.7

MS. received 1st July 1948, and in amended form 16th August 1949

ABSTRACT. Quantitative studies of extra-foveal colour vision are inherently difficult owing to the poor discrimination and rapid adaptation found in the parafoveal and peripheral areas of the retina. Measurements of the luminosity, colour mixture and hue discrimination curves have been made by a single observer using a field subtending $1^{\circ} 20'$ and located at 2° and 4° from the fovea. The results suggest that for high retinal illumination parafoveal colour vision does not differ in essential characteristics from foveal colour vision, but at low illuminations colour perception becomes reduced to a type of dichromatic vision similar to the tritanopic form of defective colour vision and to that reported for small-field foveal vision.

§ 1. INTRODUCTION

THE most exhaustive studies of colour perception have in the past been made on foveal vision using fields subtending one or two degrees. Recently interest has been directed to the study of variations in colour vision across the retina, and experiments with quite small fields, of the order of 10–15 minutes of arc, have been carried out in an attempt to analyse the colour mixture and colour discrimination functions under conditions in which a relatively small number of receptors are stimulated. Interesting results (Willmer and Wright 1945, Thomson and Wright 1946, Stiles 1949) have been reported on variations of colour sensitivity even within that area of the fovea which it is customary to regard as being free from rods, and an extension of these investigations beyond the foveal boundary into the mixed rod-cone regions of the retina is obviously desirable. From such experiments we might hope to obtain a clearer idea of the manner in which the rod and cone processes interact and also of the differences, if any, between foveal and parafoveal cone sensitivity. Moreover, under everyday conditions of seeing, quite large areas of colour are normally in view, so that a study of parafoveal colour vision should have a direct bearing on practical colour problems.

* Now at Building Research Station, Garston, Watford, Herts.

No very great reliance can be placed on experiments in which the appearance of a colour seen by parafoveal or peripheral vision is described verbally and compared with its foveal appearance. For certain purposes such comparisons may be necessary and valid, but more quantitative data are to be preferred. The aim here has been to measure the luminosity, colour mixture and hue discrimination curves for parafoveal vision by procedures essentially similar to those which have been used in the past for foveal vision.

It has proved very difficult to obtain results having a precision and accuracy sufficiently high to yield a significant measure of some of the variations of colour perception that occur across the retina. The main reasons for this appear to be the drop in visual acuity, the reduced ability to discriminate luminosity and colour differences, and the greater and more rapid adaptation in parafoveal vision compared with foveal vision. The poorer visual acuity implies that if the two halves of a colorimeter field are made too small, they will no longer be resolved in the parafovea, in which case colour matching and colour discrimination become impossible. On the other hand, if the field were made too large, the possibility of exploring individual areas of the retina would be lost, since a field subtending 6° , for example, and placed with its centre located at a point 3° to one side of the foveal fixation target, would evidently stimulate a large fraction of the fovea in addition to the parafovea. The retinal structures viewing the field would then vary continuously and markedly along the horizontal diameter of the field.

In the present experiments, a square field subtending $1^\circ 20'$ at the eye was used, this size enabling the two halves of the field to be resolved while at the same time retaining fairly local stimulation of the retina. With this size of field, however, discrimination of both luminosity and colour differences was not good, which in turn meant that precision of matching was poor. This was undoubtedly due in part to the smallness of the field and also to the inherent difficulty of making observations on a part of the field at which the observer is not looking directly.

The greater adaptability of the parafoveal retina also tends to lower the accuracy of observation, since different stimuli may give rise to rather similar sensations owing to compensatory adjustments in the sensitivity of the retinal areas used to view them. It can even happen that when prolonged and steady fixation is maintained on the fixation target, so that the parafoveal field under investigation is illuminating the same part of the retina for some time, the field may almost, if not quite, disappear. Fixation for more than a few seconds at a time has, therefore, to be avoided.

§ 2. APPARATUS AND OBSERVING CONDITIONS

The apparatus used in the experiments was the Wright colorimeter (Wright 1946). The observer's head was fixed by biting on a dental impression and was positioned so as to bring the exit-pupil of the colorimeter in line with the axis of the observer's eye. The field subtended $1^\circ 20'$ at the eye and was divided horizontally into two equal parts. The upper half, the matching field, was illuminated by any or all of three monochromatic radiations, of wavelengths 0.65μ , 0.53μ and 0.46μ , or, if required, by monochromatic light of any wavelength through the spectrum, while the lower half, the test field, was illuminated by either one or two monochromatic radiations of any desired wavelengths. A small red fixation spot, displaced horizontally from the test field, enabled the latter to be imaged on the required part of the parafovea when the fixation spot was viewed foveally. Provision was also made for the test field to be separated from the matching

field by a few degrees to permit direct comparison of the sensitivity of one part of the retina with another. The adaptation system on the colorimeter was available to condition the eye to various degrees of light and colour adaptation.

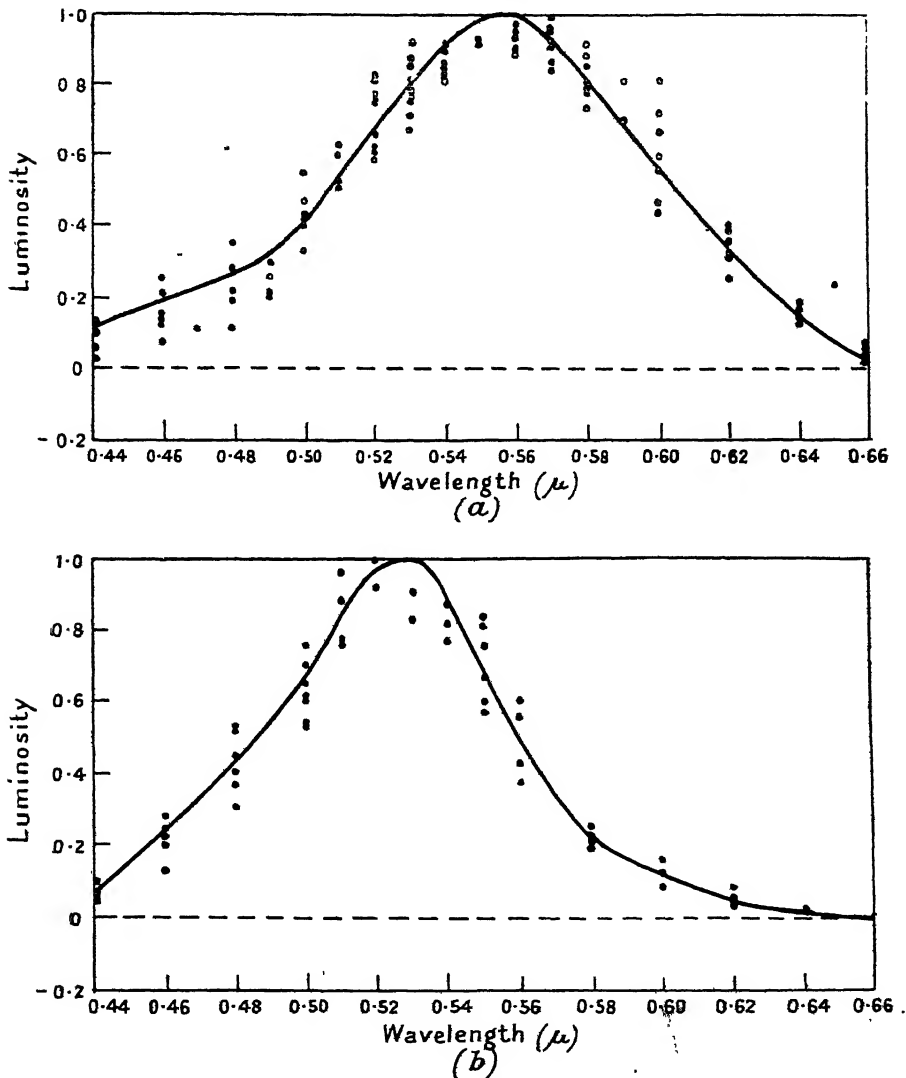


Figure 1. Luminosity curves for the parafoveal retina at a position 4° from the foveal centre, using linear luminosity scale on the ordinates. The magnitude of the errors in making the observations is indicated by the spread of the individual observed points shown in the diagram. The curves have been drawn through the arithmetic means of the observed points and the relative luminosity values have been adjusted so that the maximum of each curve is always unity.

(a) Curve measured with the eye adapted to white light (brightness of the order of 1,000 photons, colour temperature $2,848^\circ \text{K}$, approximately) and with a test field of a similar brightness.

(b) Curve measured with the eye dark adapted, and with a test field brightness of approximately 2.2 photons.

Unless otherwise indicated, the measurement and calculation of the luminosity curves, colour mixture curves and hue discrimination curves were carried out essentially as described by Wright (1946) for foveal vision.

§ 3. RESULTS

Luminosity

Typical luminosity curves for the parafoveal retina at a position 4° from the foveal centre are illustrated in Figure 1, curve (a) being recorded for the light-adapted eye and curve (b) with the eye dark-adapted and using a relatively low retinal illumination. The curves were determined by direct comparison of the test field against a monochromatic wavelength, as described by Wright (1946). Several comparison wavelengths, e.g. 0.46μ , 0.65μ , were used on different occasions for repeat observations of each curve. These curves illustrate the commencement of the well-known Purkinje shift, but in some cases humps and irregularities have been recorded, more especially in the neighbourhood of 0.51μ and 0.56μ . These have presumably had their origin in the simultaneous activity of both rods and cones under conditions tending to accentuate the sensitivity of either rods or cones at the wavelength of their maximum sensitivity. Unfortunately, the reliability of the observations has been insufficient to analyse the humps convincingly.

Colour Matching

No major differences between the foveal and parafoveal spectral coefficient curves were found at high illumination. Measurement of the white point (illuminant S_B , colour temperature $4,800^\circ \text{K.}$) showed systematic variations with distance from the fovea, as indicated by curve *a* in the chromaticity chart, Figure 2. The movement of the point towards the blue corner of the chart with

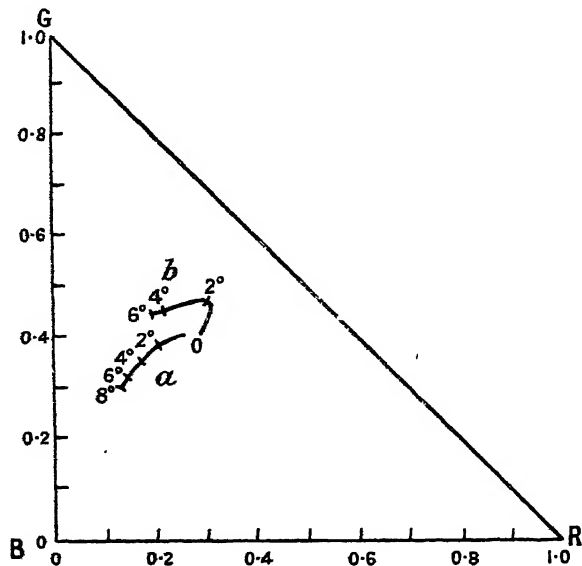


Figure 2. Chromaticity chart showing variations of the white point (standard illuminant S_B , colour temperature $4,800^\circ \text{K.}$) with distance from the fovea. Eye adapted to light of approximate colour temperature $2,848^\circ \text{K.}$, and brightness 1,000 photons. The wavelengths at the points R, G and B are 0.65μ , 0.53μ and 0.46μ respectively, and the units marked along the side of the triangle refer to the unit trichromatic coordinates. Observations were made with:

(a) Test and comparison fields both observed in the parafovea by directing the gaze at a small red fixation point. The figures 0° to 8° on the curve *a* indicate the coordinates measured with both fields at those distances from the fovea.

(b) Test field viewed in the parafovea whilst the comparison field is viewed foveally. The figures 0° to 6° on the curve indicate the coordinates measured with the test field at those distances from the comparison field at the fovea.

increasing distance from the fovea may correspond to the diminishing density of the yellow macular pigment, but this interpretation is complicated by the simultaneous variation of both retinal structure and macular pigmentation.

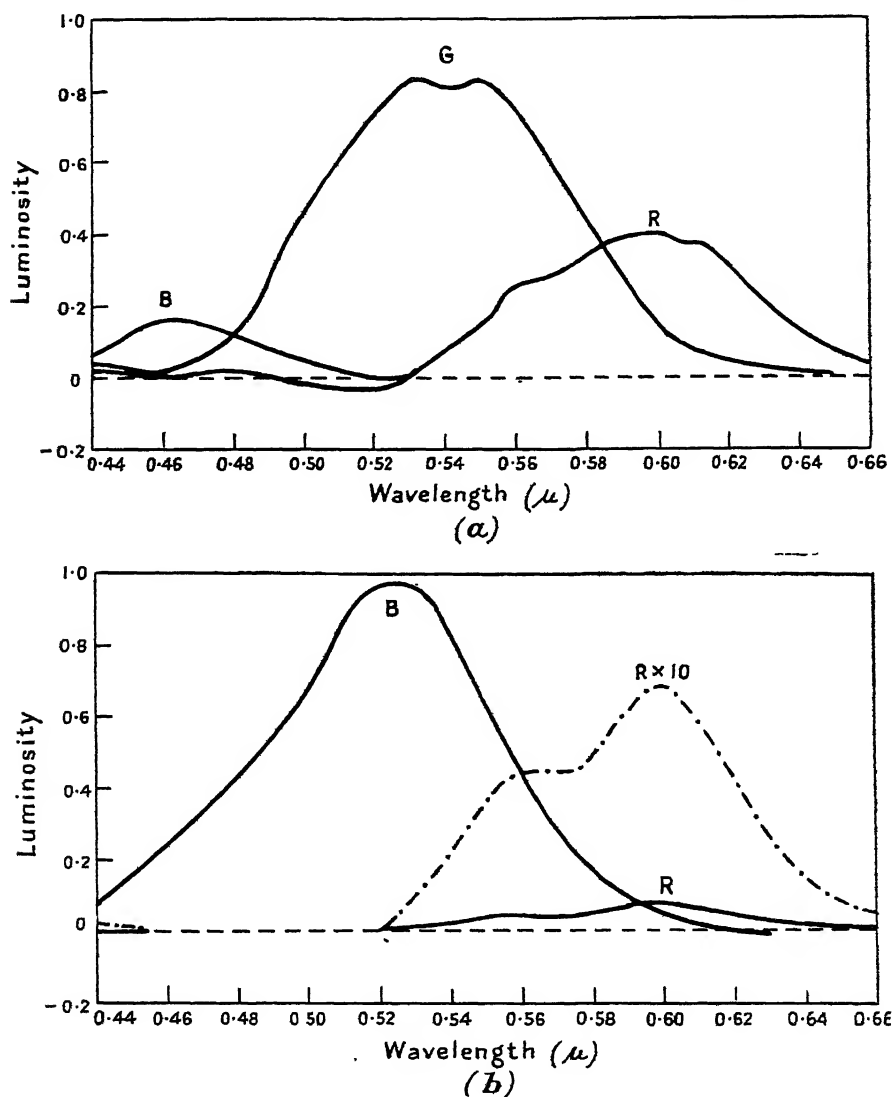


Figure 3. Colour mixture curves for the parafoveal retina at a position 4° from the foveal centre, using linear luminosity scale on the ordinates. These curves have been derived from the corresponding luminosity curves of Figure 1 and measurements of trichromatic or dichromatic coefficients, the calculations being made as described by Wright (1946).

(a) Measurements made with the eye adapted to white light (brightness of the order of 1,000 photons, colour temperature $2,848^\circ \text{K}$. approximately) and with a test field of the same order of brightness.

(b) Measurements made with the eye dark-adapted, and with a test field brightness of approximately 2.2 photons.

When colour matches were made between the test field viewed at various positions in the parafovea and the red-green-blue mixture viewed foveally, it was found that relatively less red and more green was required to match a monochromatic yellow test colour, as the test field was displaced parafoveally.

This difference can only be due to differences in the visual processes and not to pigmentary absorption, since the latter cannot cause any change in the quality of monochromatic light. The corresponding effect on the white point is shown by curve *b* in Figure 2.

At low illuminations and with both fields viewed parafoveally, colour matching tended to become dichromatic. All the spectral colours, for example, could be matched by a mixture of only two stimuli, one of these being a red radiation (for example 0.65μ) and the other either a green, blue-green or blue (for example 0.46μ). The dichromatic mixture curves obtained at 4° displacement into the parafovea are illustrated in Figure 3 (*b*). Two points of special interest are the small contribution made by the red mixture curve to the total luminosity curve, and the hump in the red curve itself. There is such a striking similarity to the red mixture curve found by Willmer and Wright (1945) for small field foveal vision, that the hump must almost certainly have a common origin in the two cases. The trichromatic mixture curves obtained at 4° displacement into the parafovea when the eye is light adapted are illustrated in Figure 3 (*a*). Both the red and green curves show humps in this case.

Hue Discrimination

The most noticeable difference between foveal and parafoveal hue discrimination at high illuminations was the larger wavelength difference required in parafoveal vision to produce a just noticeable difference in hue. A typical curve

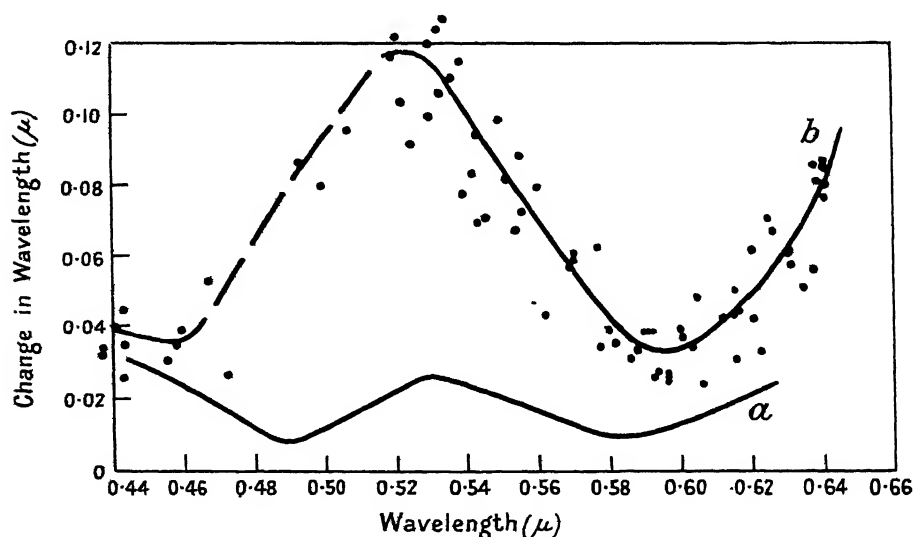


Figure 4. Hue discrimination curves for the parafoveal retina at 4° from the foveal centre. Individual observed points give an indication of the errors of observation involved.

(*a*) Eye adapted to light of brightness approximately 1,000 photons, colour temperature $2,848^\circ \text{K}$, and test field of the same order of brightness.

(*b*) Eye dark-adapted and test field brightness approximately 1.1 photon.

for 4° parafoveal viewing is shown in Figure 4, curve *a*, in which maximum sensitivity (minimum wavelength difference) occurs in the orange and blue-green parts of the spectrum, as with foveal vision.

At low illuminations, however, there is a definite alteration in the type of discrimination curve, since discrimination in the blue-green wavelengths becomes

very poor indeed. Curve *b* of Figure 4 consists of a single minimum in the orange, with some isolated points in the blue and violet. There is, in effect, a discontinuity in the blue-green, similar to that reported by Willmer and Wright (1945) for small field foveal vision, since the blue and green wavelengths are virtually indistinguishable.

§ 4. CONCLUSIONS

Many of the experimental data, more especially those relating to the effect of adaptation of parafoveal colour vision, have been omitted from this report because of lack of precision of the measurements. The first conclusion which must be drawn therefore is the inherent difficulty of securing results on parafoveal colour vision which can have any claim to useful accuracy. From the more conclusive of the results which have been recorded it appears (i) that parafoveal colour vision, at high illuminations does not differ in essential characteristics from foveal vision, although secondary differences do exist, and (ii) that at low illuminations, colour discrimination becomes less acute and is reduced to a form of dichromatic vision which is rather similar to the tritanopic class of defective colour vision and also to that reported for small field foveal vision.

One explanation given for this type of reduced discrimination is inactivity of the blue receptors. Alternatively, it may be that the presence of rods in the parafovea adds a colourless quality to the blue and blue-green response at low illuminations, which swamps the blueness and greenness, whereas at the orange and red end of the spectrum, the rods are relatively much less sensitive than the cones and little or no dilution of hue can occur.

ACKNOWLEDGMENTS

This work was carried out as part of a research programme under the direction of Professor W. D. Wright; his constant help and advice during the observations and in preparing this paper are greatly appreciated.

My thanks also go to Misses M. E. Radley and M. J. Thornley, Dr. L. C. Thomson and Mr. J. H. Dodwell for help in recording the observations, and to the Medical Research Council for their financial support of the investigation.

REFERENCES

- STILES, W. S., 1949, *Rev. d'Optique*, **28**, 215.
THOMSON, L. C., and WRIGHT, W. D., 1946, *J. Physiol.*, **105**, 316.
WILLMER, E. N., and WRIGHT, W. D., 1945, *Nature, Lond.*, **156**, 119.
WRIGHT, W. D., 1946, *Researches in Normal and Defective Colour Vision* (London: Kimpton).

Quantum Efficiency in Photographic X-ray Exposures*

D. BROMLEY† AND R. H. HERZ

Research Laboratories, Kodak Ltd., Wealdstone, Harrow, Middlesex

MS. received 30th May 1949, and in amended form 4th August 1949

ABSTRACT. The grain yield per absorbed quantum has been evaluated experimentally for four different photographic emulsions over a wavelength range of $1.2\text{--}0.01\text{ \AA}$, i.e. from very soft x rays to gamma rays emitted from radium. The tests were carried out with both heterogeneous and monochromatic x rays. A grain yield of 1.0 per absorbed quantum appears to hold within a range of wavelengths of $1.2\text{--}0.37\text{ \AA}$. (10 kv.–33 kv.). From 0.37 \AA to short wavelengths (gamma rays) an increase of grain yield linear with quantum energy is obtained. At a wavelength of 0.0123 \AA . (1,000 kv.) about 80 grains are rendered developable per quantum absorbed. The rate of increase of grain yield with quantum energy is different for different emulsions. For a slow and fine-grain chloro-bromide emulsion the grain yield was found to be less than unity for wavelengths below 0.12 \AA . (i.e. 100 kv.). The results are consistent with the hypothesis that grains are made developable not only by the direct action of x rays but also by secondary electrons produced by the x rays. These are effective only when they are energetic enough to reach grains in the neighbourhood of that absorbing the original x-ray quantum.

§ 1. INTRODUCTION

IN order to understand the mechanism of direct photographic x-ray exposures a knowledge of the efficiency of the x-ray energy absorbed in photographic emulsions as a function of the wavelength is required. This efficiency may be expressed as the number of silver bromide grains rendered developable per absorbed quantum, often called the quantum efficiency or the grain yield per absorbed quantum.

Eggert and Noddack (1927) determined the grain yield per absorbed quantum of x rays in photographic emulsions and found that each quantum absorbed renders approximately one silver bromide grain developable. They used a heterogeneous radiation equivalent to a monochromatic radiation of wavelength 0.45 \AA . The determination of this equivalent wavelength was based on calorimetric energy measurements. Silberstein and Trivelli (1930) gave somewhat less direct evidence that the one quantum hypothesis holds in the range of x-ray wavelengths. Similar investigations to those by Eggert and Noddack were carried out by Günther and Tittel (1933) within the range of wavelengths from 0.245 \AA . to 1.54 \AA ., which agreed well with the results of the former workers. In addition they found that the number of silver atoms set free is proportional to the energy of the x-ray quantum. Pelc (1945) has recently given a theoretical and experimental account of the increase of grain yield with increasing quantum energy.

The purpose of the work described in this paper is to examine the quantum efficiency of various types of photographic emulsions over a wide range of x-ray quality, by providing special experimental techniques in evaluating the grain yield. These techniques consisted in measuring the actual absorption of x rays by emulsions stripped from film support, and in using in some cases monochromatic x rays instead of heterogeneous radiation. By this latter procedure a greater absolute accuracy in the determination of the grain yield per absorbed quantum was expected.

§ 2. EXPERIMENTAL METHODS AND CALCULATIONS

Experiments were carried out in the first place by exposing films to measured doses of filtered heterogeneous x rays whose quality was determined in terms of half-value layers in copper by means of an ionometric half-value layer meter. The half-value layer values were converted into absorption coefficients and the equivalent wavelength of the radiation was then determined as the wavelength of that monochromatic radiation having the same absorption coefficient.

In four cases monochromatic x rays were used from secondary radiators, such as platinum, copper and barium carbonate, which acted as characteristic x-ray sources when irradiated by heterogeneous x rays from a tungsten target tube. The K_β radiations were suitably filtered by selective absorption filters. This method of obtaining monochromatic x rays was regarded as preferable to the use of a monochromator because it provides a fairly wide and uniform beam, suitable for density and ionization measurements.

The dose was measured by means of a Victoreen condenser r-meter, or, in the case of very soft x rays, by means of another ionization dosimeter using an ionization chamber with 'Cellophane' window. Both meters were calibrated against a free air standard ionization chamber at the National Physical Laboratory. The number of x-ray quanta, Q_i , incident upon the emulsion for various wavelengths was computed from the definition of the roentgen dose. The number of quanta absorbed (Q_a) per unit area in the photographic emulsion was then calculated as $Q_i(\mu/\rho_{\text{AgBr}})w$, where μ/ρ_{AgBr} is the mass absorption coefficient of silver bromide, and w the weight of the silver bromide in gm/cm^2 of emulsion. Details of this computation are described below.

The number of developed grains in the emulsion was then counted under the microscope, using a squared graticule and utilizing only that emulsion layer (in the case of double coated x-ray films) facing the x-ray tube during exposure. The number of fog grains, counted on unexposed and developed film strips, was subtracted from the former number. Since the counting of grains under the microscope is tedious and the statistical error involved rather great, the number of grains rendered developable was also evaluated from density measurements, assuming that density is proportional to the number of grains affected. These tests were limited to the range where the relationship between density and exposure is linear.

The desired number of developed grains per quantum absorbed in the emulsion was found from the ratio of the number of grains counted per unit area to the number of quanta absorbed per unit area.

The energy of x rays absorbed E_{abs} within a small mass of air dm may be calculated from the product of the energy flux E_{fl} in ergs/cm^2 , the mass in grammes of the air, and the fraction of the energy flux, incident upon the mass, that is converted into kinetic energy of electrons. This fraction is given by $(\tau + \sigma_a)_{\text{air}}$, where τ is the photoelectric absorption coefficient and σ_a the true absorption coefficient associated with Compton recoil electrons. Thus

$$E_{\text{abs}} = E_{\text{fl}} dm(\tau + \sigma_a)_{\text{air}} \quad \dots\dots(1)$$

Hence

$$E_{\text{fl}} = \frac{E_{\text{abs}}}{dm(\tau + \sigma_a)_{\text{air}}} \quad \dots\dots(1a)$$

By definition, one roentgen unit (r-unit) produces 1 E.S.U. of ions of each sign in 1 cm^3 of air. The number of electrons collected in $1 \text{ cm}^3 (= 0.001293 \text{ gm.})$ of air per roentgen is $1/e$, where e is the electronic charge in E.S.U. The energy

required to produce one pair of ions in air is approximately 33 electron volts. Thus the energy absorbed per cm^3 of air for 1 roentgen dose is $33/e$ electron volts. Since $1 \text{ ev.} = e/300$ ergs, the dose of 1 roentgen corresponds to the absorption of $33/300 = 0.11$ ergs. Since 1 roentgen corresponds to the absorption of 0.11 erg in 1 cm^3 of air, equation (1a) may be written

$$E'_{\text{fl}} = \frac{0.11}{dm(\tau + \sigma_a)_{\text{air}}} \text{ erg/cm}^2, \quad \dots\dots(1b)$$

where E'_{fl} is now the energy flux per roentgen. This expression is a function of wavelength through the term $(\tau + \sigma_a)_{\text{air}}$. The variation with wavelength of the energy flux to give 1 roentgen in air has been computed and is shown graphically in Figure 1. In accordance with similar computations by Mayneord (1940) the curve possesses a maximum at about 0.1 \AA . where the energy flux is more than six times as great as at 0.4 \AA .

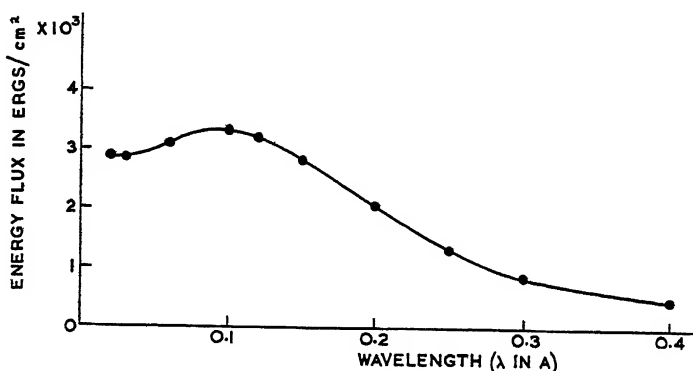


Figure 1. Relationship between energy flux in ergs/cm² and wavelength for 1 r-unit.

The energy absorbed, at a given wavelength, per cm^2 of a thin layer of silver bromide emulsion can be computed for any dose by multiplying expression (1b) by the mass absorption coefficient of silver bromide, by w , the coating weight per cm^2 of the emulsion, and by the number of roentgens, r . Thus

$$E_{\text{abs}} = \frac{0.11\mu w r}{(\tau + \sigma_a)_{\text{air}} \rho} \text{ erg/cm}^2. \quad \dots\dots(2)$$

The gelatin absorption is neglected in these considerations as calculations have shown that the error involved is less than 1%. The energy flux E_{fl} for one roentgen unit may be converted into the number of quanta Q by dividing it by the energy of the quantum concerned, thus

$$Q = \frac{E_{\text{fl}}}{h\nu} = \frac{\lambda E_{\text{fl}}}{hc} = \frac{0.11\lambda}{hc(\tau + \sigma_a)_{\text{air}}} = \frac{\lambda}{(\tau + \sigma_a)_{\text{air}}} 5.53 \times 10^{14}.$$

If λ is expressed in Ångström units

$$Q = \frac{\lambda}{(\tau + \sigma_a)_{\text{air}}} 5.53 \times 10^6. \quad \dots\dots(3)$$

The number of quanta absorbed per roentgen unit in 1 square centimetre of a silver bromide emulsion is then obtained by multiplying the expression in equation (3) by the combined mass absorption coefficient of silver bromide and

the weight of the silver bromide per cm² of emulsion (neglecting the gelatin). If the density of AgBr is taken as 6.47, then

$$Q_{\text{abs}} = \frac{\mu w \lambda}{(\tau + \sigma_a)_{\text{air}}} 0.855 \times 10^6. \quad \dots\dots (4)$$

The combined mass absorption coefficient of silver bromide for various wavelengths was found according to

$$\mu_{\text{AgBr}}/\rho = \frac{w_1}{w_1 + w_2} \cdot \frac{\mu}{\rho_{\text{Ag}}} + \frac{w_2}{w_1 + w_2} \cdot \frac{\mu}{\rho_{\text{Br}}},$$

where w_1 and w_2 are the weights of silver and bromine respectively.

§ 3. SELECTION AND DETERMINATION OF WAVELENGTHS

(i) *Heterogeneous Radiation*

In order to cover a wide range of wavelengths in the first instance heterogeneous radiations were used. The generating kilovoltage, filtration and the equivalent wavelengths of these radiations are given in Table 1. The equivalent wavelengths were derived from half-value layer measurements partly on copper and partly on silver.

Table 1

kv.	Filtration	Equivalent λ (A.)
50	0.4 mm. Cu	0.30
80	0.8 mm. Cu	0.21
120	1.2 mm. Cu	0.16
195	4.0 mm. Cu	0.10
475	4.53 mm. Sn	0.047
1000	2 mm. Al + 4.2 mm. Fe + 6.5 mm. Pb	0.024
Gamma rays	1.0 cm. Pb	0.0105

The results on quantum efficiency which are given in § 5, are based on the wavelengths tabulated above.

As there was some doubt in regard to the accuracy of the wavelength determination of heterogeneous radiations which might seriously affect the results on quantum efficiency, the wavelengths were checked indirectly by absorption measurements with stripped silver bromide emulsions. The results of these were then compared with those obtained from computation using the mass absorption coefficients for the wavelengths from half-value layer measurements.

The absorption measurements were made by means of an x-ray ionization dosimeter comprising a Lindemann electrometer and two thimble ionization chambers of identical design and equal volume. The central electrodes of the two thimble chambers were connected to the needle of the electrometer. The walls of the chamber were connected to equal and opposite potentials (54 v.). If the volumes of the chambers are equal and if the chambers are exposed to a common uniform x-ray beam, so that both chambers receive equal doses of x rays, the electrometer needle will be in electrical balance. If a stripped photographic emulsion having no support is placed above one of the chambers the electrometer needle will be out of balance by an amount proportional to the absorption in the film which may thus be measured. The percentage absorption is given by $100(I_0 - I_1)/I_0$, where I_0 is the intensity of incident radiation, and I_1 the intensity transmitted by the emulsion. The method described measures $I_0 - I_1$ directly, leaving the separate values of I_0 and I_1 unknown. The value of

I_0 was determined simultaneously by means of the Victoreen condenser r-meter. The method is thus substantially independent of fluctuations in the beam, whereas in measuring such small absorptions by successive readings with and without absorber, errors in the meter readings and intensity differences due to fluctuations of the intensity of the beam are generally of the same order as those due to absorption. In actual use without the absorber in position, an exact balance is not easily obtained, but a small residual drift remains. This drift, with appropriate sign, was therefore subtracted from the deflection due to absorption as actually measured.

Absorption measurements were made with 1, 2, 3 . . . superimposed stripped emulsions placed about 1 inch above one of the chambers. The results obtained, which were plotted as percentage absorption against the number of stripped emulsions, showed more than a proportional increase in absorption with the number of sheets. The effect was ascribed to scatter and fluorescent x rays from the absorbing sheets. The influence of scatter was however eliminated by placing the absorbing sheets 8 in. above the measuring chamber.

A typical result of such an absorption measurement using a stripped emulsion without base is shown in Figure 2. The absorption of one sheet may be read

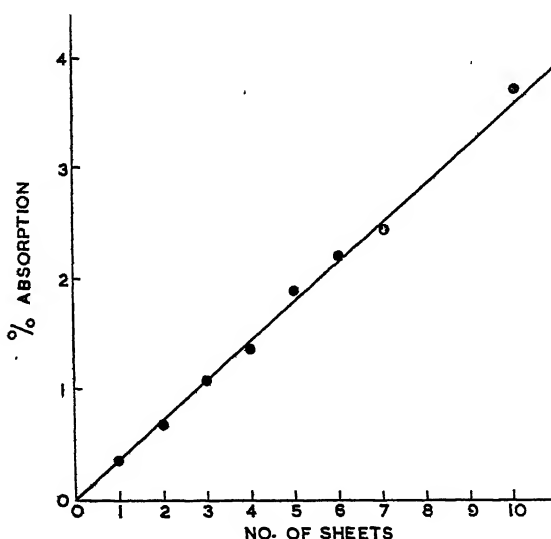


Figure 2. Relationship between percentage absorption and number of stripped emulsion layers for x-rays generated at 80 kv., 8 ma., 0.8 mm. Cu-filter (equivalent wavelength $\lambda=0.21 \text{ \AA}$.)

from the graph and is 0.37% for radiation the wavelength of which was computed from half-value layer measurements to be 0.21 \AA . The percentage absorption to which this wavelength leads using the absorption coefficients from tables is in this particular case exactly the same. It may be mentioned that the film used was a chloro-bromide emulsion and that the constituents of this emulsion were considered in calculating the absorption for various wavelengths. Absorption measurements were also carried out on another type of stripped emulsion using radiations generated at 50 kv. filtered by 0.4 mm. Cu. at 80 kv., (0.8 mm. Cu), and at 120 kv. (1.2 mm. Cu).

In addition half-value layer measurements of these radiations in silver and copper were made and, using the equivalent wavelengths, the percentage absorption was calculated. The calculated and the experimentally obtained percentage

absorption values agree fairly satisfactorily. The greatest deviation occurring between two differently evaluated wavelengths was not more than 7%.

It was mentioned above that half-value layer figures were also determined in terms of millimetres of silver. This was done as it was found that the effective wavelength of the same heterogeneous radiation differed in different material. Since the absorption in the silver bromide emulsion is predominantly that of silver, it was thought that the half-value layer in silver would give a better estimate of the actual quality of x rays, so far as the emulsion is concerned. The results show however only small differences between the half-value layers obtained in copper and in silver.

(ii) *Monochromatic Radiation*

In view of the uncertainty of the wavelengths used as computed from heterogeneous radiations, experiments were carried out with the object of producing monochromatic radiations. The methods available are (1) reflection from a crystal, (2) use of secondary characteristic radiation excited by the absorption of primary x rays. Method (1) gives a sufficiently intense but narrow beam of x rays although any particular line of known wavelength may be selected or any wavelength may be calculated using the Bragg relation. Ionization measurements in terms of r-units become difficult, however, because of the narrow width of the

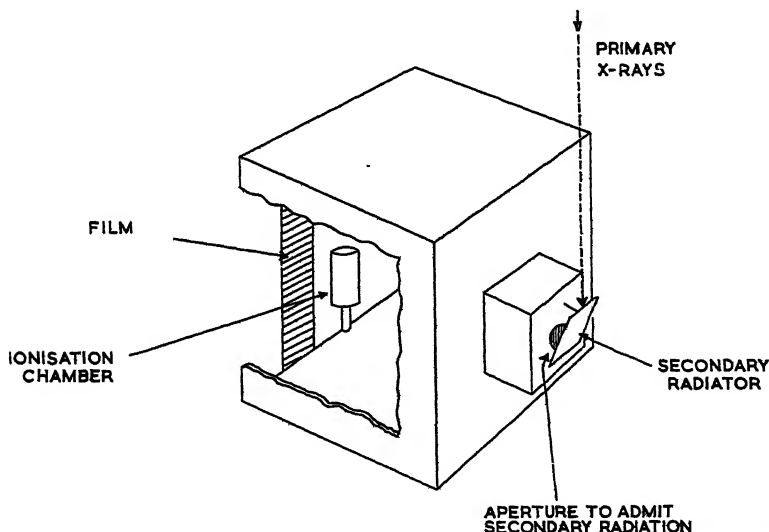


Figure 3. Arrangement for generating monochromatic x-radiation and simultaneous measurement by film and ionization chamber.

beam. Method (2) gives a broad beam of low intensity. Careful screening from the primary beam is necessary. The spectrum normally comprises several lines together with scattered and continuous radiation. The latter method appeared to be preferable for this investigation chiefly because preliminary tests have shown that a wide and relatively uniform x-ray beam could be obtained, which allowed film exposure and ionization measurements to be made simultaneously*. The experimental set up is illustrated in Figure 3. The apparatus consists mainly of a large lead lined box. One side of this is removed, and to the side

* A similar method adopted at about the same time in the Research Laboratories of Eastman Kodak Company, U.S.A., for sensitometry (private communication by H. E. Seemann and M. H. Horn) is interesting because, although known before the establishment of crystal reflection, the method has rarely been used.

directly opposite is fixed a smaller box with an aperture for admitting secondary radiation. The ionization chamber and the film are placed in the open side of the box opposite the aperture. The primary radiation impinges on a sheet of the secondary radiator (platinum, copper or a block of barium carbonate). The main purpose of the box is to prevent primary x rays from passing on to the film and to the ionization chamber, which were placed parallel to the plane of the aperture at a distance of 10 inches from the latter, as indicated in Figure 3. The characteristic radiations used for this investigation and the selective absorption filters to eliminate scatter and K_β radiations are given in Table 2.

Table 2

Wavelength (A.)	Radiation	Filter
0.188	K_α platinum	Selective filter CaWO_4 .
0.39	K_α barium	Selective filter KI.
1.2	L platinum	No filter, excited by x rays generated at 70 kv., i.e. below excitation potential of K lines of platinum.
1.4	K_α copper	Selective absorption filter Ni.

These wavelengths were selected because they cover a fairly wide range and because the secondary radiators were easily obtainable.

The selective absorption filters were prepared as follows :

For the K_α platinum line a Kodak Ultra-Speed back intensifying screen was used. For the K_α barium line small crystals of potassium iodide were mixed with cellulose lacquer coated on card and placed between two sheets of x-ray film base. The weight of iodine per cm^2 required for the absorption of the K_β line was calculated on the basis of a transmission of 1/200 of the incident intensity. For the K_α copper radiation a nickel foil of 0.002 in. was used.

In order to check whether the radiations used were really monochromatic, a crystal spectrograph was designed for the purpose. The spectrograph was used with a stationary crystal as the crystal received radiation from different angles since the radiation from the secondary radiator was diffused. Spectrograms of the K lines of barium and platinum and the L lines of platinum show that the K_β line of barium was selectively absorbed by KI. In the case of platinum the K_α and K_β lines are not separately resolved, but it can be assumed that the K_β line was partly eliminated by the selective absorption of a sheet of calcium tungstate (CaWO_4). No background density due to scattered or continuous radiation is evident in the spectrograms.

A further check on the quality of the K radiation of barium and platinum was made by means of photographic absorption measurements using the method described by Seemann and MacGillivray (1946). The absorption curves obtained were almost linear. It was not possible to check whether the small deviation from linearity is due to a small background effect or to inaccuracy of the photographic method. The absorption coefficients of the two radiations were computed from the mean slopes of the absorption curves, and the equivalent wavelengths are set out in Table 3.

It is not possible to distinguish between the K_{α_1} and K_{α_2} lines, but the difference between them is in any case negligible. The wavelength used was therefore a mean of the K_{α_1} and K_{α_2} wavelengths weighted in the ratio 2:1 according to their relative intensities.

The quality of the platinum L radiation was checked by absorption measurements using stripped x-ray emulsions and also aluminium sheets of 0.1 mm. thickness. These absorption measurements were made by means of a 'grenz-ray' ionization chamber, which is provided with a 'Cellophane' window. The percentage absorption of one sheet of stripped emulsion derived from the absorption curve is 16% as against 17.8% when calculated from the known wavelength of platinum L radiation. The discrepancy between the two values may well be caused by the uncertainty in the mass absorption coefficient of silver bromide in this range of wavelengths. The mass absorption coefficient on which this calculation is based is 88.5.

In the case of the L radiation of platinum the effective wavelength was computed from a mean of the strongest L_{α} and L_{β} lines, from which a value of 1.2 Å. was derived.

Table 3

Radiation	Wavelength (Å.)	Wavelengths (Å.) evaluated from photographic absorption curves
K_{α} platinum	0.178	0.183
K_{α} barium	0.39	0.37

Table 4. Monochromatic Radiation

Variation of density over film area

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
A	0.47	0.54	0.58	0.61	0.62	0.61	0.61	0.60	0.58	0.55
B	0.57	0.61	0.64	0.66	0.65	0.65	0.64	0.63	0.61	0.60
C	0.58	0.63	0.66	0.68	0.70	0.69	0.68	0.67	0.65	0.65
D	0.59	0.66	0.69	0.73	0.74	0.74	0.72	0.71	0.68	0.67
E	0.61	0.66	0.72	0.75	0.78	0.79	0.77	0.75	0.73	0.71
F	0.58	0.64	0.70	0.73	0.76	0.77	0.76	0.75	0.72	0.71
G	0.59	0.63	0.68	0.73	0.76	0.77	0.75	0.75	0.71	0.68
H	0.57	0.63	0.68	0.72	0.75	0.76	0.75	0.73	0.69	0.65

No spectrogram was obtained of the CuK_{α} line owing to temporary difficulties with the spectrograph which had not been overcome when this paper was written. A check on the secondary radiation of Cu was made by means of ionometric absorption measurements, which showed that the radiation must have been slightly contaminated. It is probable therefore that the results on quantum efficiency using CuK_{α} radiation are less reliable than those obtained with the other radiations.

An essential test in regard to the uniformity of the diffused characteristic radiation thus obtained had to be carried out, since ionization chamber measurements and film strip exposures had to be made side by side simultaneously. The uniformity was examined by exposing to the monochromatic radiations a sheet of film which covered the positions of the plane of the film and the plane through the centre of the ionization chamber. The result obtained on the test film is shown in terms of densities measured at intervals over the film area of 6 in. \times 8 in. in Table 4. The density values framed on the left of the Table cover the area in which the film strip was placed. The density values framed on the right represent the measures obtained in the plane covered by the centre of the ionization

chamber. It will be seen that the mean of the densities obtained on the left and on the right agree fairly well, namely $D=0.68$ on the left and 0.72 on the right. As the centre of the chamber was placed flush with the plane of the film, the assumption was made that the dosimeter reading corresponds to the dose impinging on the film.

It may be further noted that the ionization measurements for the Pt_L and $Cu K_\alpha$ radiation, which are very soft, were made with the 'grenz-ray' chamber mentioned above. This was specially calibrated by the National Physical Laboratory for Pt_L radiation, using the same apparatus for producing the Pt_L radiation as is described above. The calibration was made against a free air standard ionization chamber.

§4. GRAIN COUNTING AND DENSITY MEASUREMENTS

The following types of emulsions were used for the evaluation of quantum efficiency: Type A, non-screen x-ray film; Type B, screen x-ray film; Type C, non-screen x-ray film of fine grain; Type D, chloro-bromide stripping emulsion. All films were processed with constant agitation in D19b x-ray developer at $68^\circ F.$ for 4 minutes in a processing machine in which the reproducibility of development is of high order.

In the first instance, very small x-ray doses were given so that grain counts under the microscope could be made as described in §2. The number of grains per unit area was then plotted against the densities, which were read on a photoelectric densitometer of a high order of reproducibility. Typical plots which illustrate the relationship between the number of grains counted per unit area and the density are shown in Figure 4. The curves cut the ordinate at zero number of grains at a finite density corresponding to the density of the base. Although it was anticipated that the quality of radiation would have no influence on the relationship between the number of grains and density, a check was made by applying various qualities of radiation. The various points refer to radiations of different wavelengths, and are indicated by various symbols, as explained on the graphs of Figure 4. The random deviation of the points indicates that no influence of the quality of x rays could be observed. The regression lines relating the number of grains per unit area against density were evaluated by correlation methods. The standard deviation of points about regression lines is about 1 grain.

§5. EVALUATION AND RESULTS ON QUANTUM EFFICIENCY

The results of the measurements described above and plotted in Figure 4 were used for the evaluation of quantum efficiency. From the known value of the dose, measured in r-units for each exposure, the number of quanta absorbed per unit area in the various emulsions were computed, using equation (4) of §2, and the quantum efficiency was computed from the ratio:

$$\frac{\text{number of developed grains per unit area}}{\text{number of quanta absorbed per unit area}}$$

Initially small x-ray doses below 0.01 r-units were given for the various wavelengths so that the number of developed grains could be counted. Since the doses were close to the limit of sensitivity of the dosimeter, greater accuracy was expected by using higher doses. In order to correlate higher dose measurements with the number of developed grains, the grain-density curves, Figure 4,

were extrapolated to higher density values, and the number of developed grains was evaluated from density readings. The justification for the extrapolation of the grain number-density curve is given by the relationship derived by Nutting (1913),

$$D = 0.4343 \bar{a} n,$$

where D = density, \bar{a} = the average projected area per grain and n = the number of developed grains.

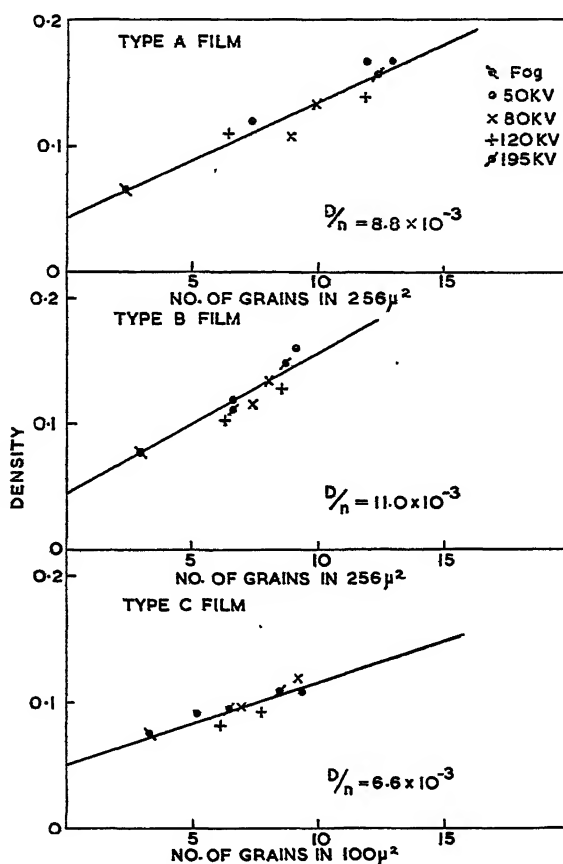


Figure 4. Relationship between photographic density and number of developed grains counted in Type A, B and C film for exposure to x rays generated at 50, 80, 120 and 195 kv.

Assuming that \bar{a} remains constant as the exposure increases for densities below $D=1$, the equation indicates proportionality between D and n . The constants of proportionality D/n of the straight line (Figure 4) were evaluated and are shown below the curves. The quantum efficiency was then computed from

$$\frac{n}{Q} = \frac{D n/D}{r Q/r},$$

where n = number of developed grains per unit area above the fog, Q = number of quanta absorbed per unit area of single emulsion, r = number of roentgen units, and D = density above fog.

The ratios D/r and n/D were obtained experimentally as described above, and the ratio Q/r was computed from equation (4) §2. The values of quantum efficiency obtained for the various frequencies of radiation and types of films are shown in the graphs (Figures 5-8) illustrating the grain yield per absorbed quantum η against the frequency ν of the radiation. Figure 5 shows this relationship for film of Types A, B and C for x rays over a range of wavelengths of

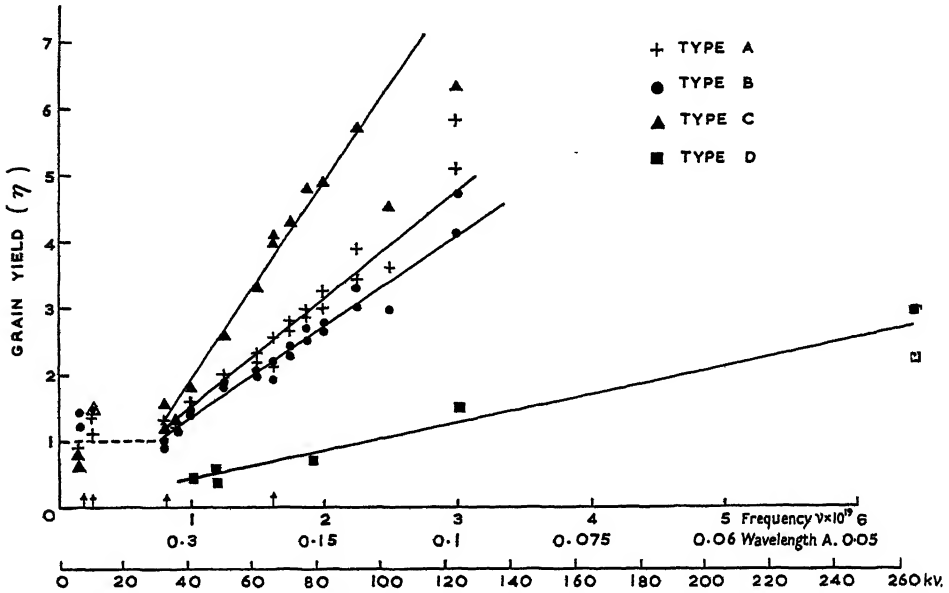


Figure 5. Relationship between grain yield and frequency of x rays for Type A, B, C and D film (exposures to monochromatic x rays are indicated on the abscissa axis by arrows).

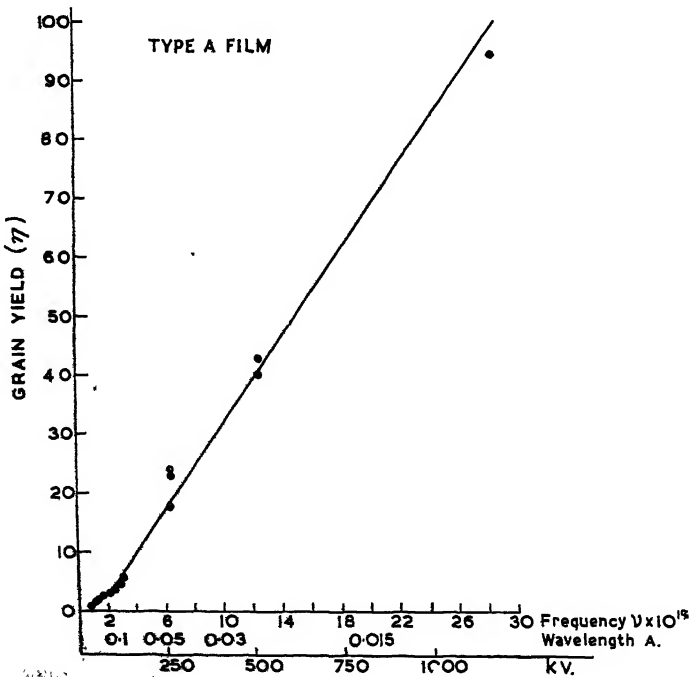


Figure 6

1.54-0.1 Å. and for Type D film over a range of 0.3-0.047 Å. For Type D film the results have been obtained from actual grain counts, whereas the remaining results were derived from the extrapolated grain number-density curves (Figure 4).

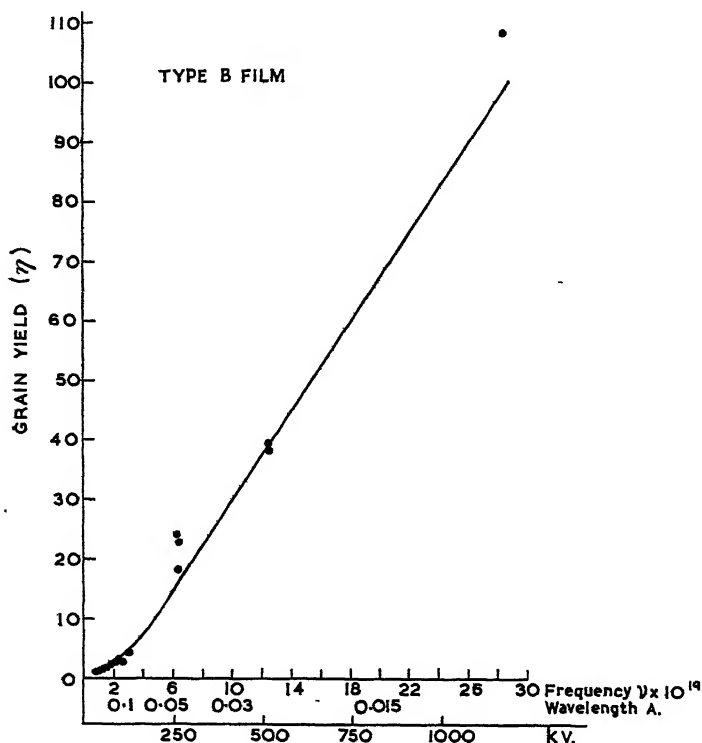


Figure 7.

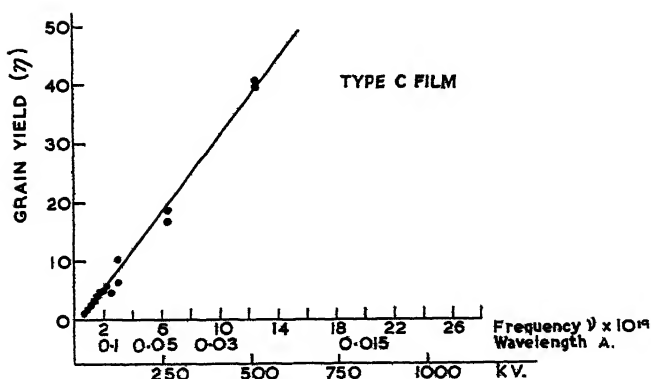


Figure 8.

Figures 6-8. Relationship between grain yield and frequency of x rays generated between 40 and 1,250 kv. (Type A, B and C film).

In order to cover also the results obtained with very hard radiation such as is generated at 475 kv. and 1,000 kv. and with gamma rays from a radium source, graphs are presented in Figures 6, 7 and 8, Types A, B and C film respectively.

It should be noted that the mean wavelength corresponding to gamma rays emitted from a radium source filtered by 1 cm. of lead is quoted as 0.0105 A., which is equivalent to 1,230 kv.

An interpretation of the results shown in Figures 5-8 is dealt with briefly in the following section and in more detail in §7.

§6. QUANTUM EFFICIENCY IN WATER-SWOLLEN FILMS AND CONSIDERATIONS OF SUB-IMAGE FORMATION

(i) *Quantum Efficiency in Water-Swollen Films*

The experimental result that several grains may be rendered developable per quantum absorbed can be understood on the assumption that the energy of the absorbed quantum is sufficient to release photoelectrons from the atoms of the affected grain which, during their passage through the emulsion, render other neighbouring grains developable. Some evidence for this effect was given by the exposure of Type A film in the wet state, where it can be assumed that the grains are further apart than in the dry state. In the wet state photoelectrons emitted from an affected grain are more readily absorbed by the water-swollen gelatin, and will be less able to reach neighbouring grains, also because of the greater average separation between grains. The quantum efficiency found for the film exposed in the wet state at a wavelength of 0.1 A. was found to be almost half of that found if the same emulsion was exposed in the dry state. This result seems to indicate that the decrease of grain yield is caused by the reduced length of path of photoelectrons, the water-swollen emulsion and the greater separation between grains (Herz 1949 a).

In the region of grain yield 'unity' it may be assumed that the path length of photoelectrons released from the affected grain is not sufficient to reach neighbouring grains, or at least that these photoelectrons do not contribute to the developability of neighbouring grains. Evidence for this assumption was given by exposure of Type A film in the wet state at $\lambda = 0.37$ A., where the grain yield is unity in the dry state of the film. The result showed that the same grain yield of unity was obtained in the wet state as in the dry state. It can be thus concluded that a grain yield of unity is obtained by single quantum hits without contributions of photoelectrons from neighbouring grains.

(ii) *Considerations of Sub-Image Formation*

It was shown by Burton and Berg (1946) that photographic emulsions showing a toe in a density-exposure curve are likely to show sub-image formation. The only emulsion used in these tests which showed a slight indication of a toe in the density-exposure curve was Type D film, which also proved to have a quantum efficiency below unity for wavelengths longer than 0.13 A. A low-intensity post-exposure of this film to light gave evidence for sub-image formation which could not be found in any of the other emulsions.

§7. INTERPRETATION AND CONCLUSIONS

From a study of Figures 5-8 the following conclusions can be drawn:

(i) *Grain yield between 1.2-0.37 A.* (Figure 5). In Figure 5, a flat part of the curves at a quantum efficiency of 1 is seen in the region of very soft radiation corresponding to a wavelength range between 1.54-0.37 A. The energy of one absorbed quantum is evidently sufficient to render one grain developable

over this relatively wide range of wavelengths in each case of the three emulsions Types A, B and C. This means that the number of ionizations produced from one quantum absorbed occurring at the shorter wavelength 0.37 Å. is more than is required to render one grain developable. The range of photoelectrons, however, appears to be insufficient for them to reach neighbouring grains. Thus, no contribution to the developability of neighbouring grains is made by one quantum hit. This supposition is further supported by the experimental result described in the last section, which showed that the quantum efficiency of the water-swollen emulsion at 0.37 Å. remains at unity.

(ii) *Grain yield between 0.37–0.0105 Å.* (Figures 6–8). The most characteristic feature of the graphs appears to be that the grain yield is proportional to the quantum energy within wavelengths from 0.37–0.1 Å. and for the Type C film down to a wavelength of 0.047 Å. The curves (Figures 6–8) do not pass through the origin, and those for films Types A and B bend at a wavelength about 0.1 Å.

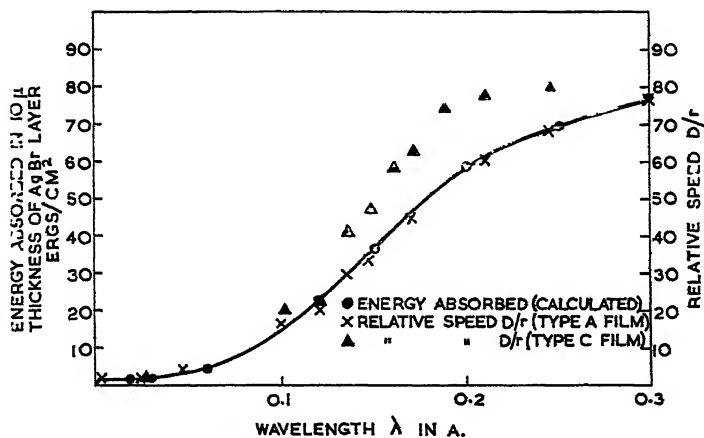


Figure 9. Relationship between energy absorbed (erg/cm²) in 10 μ layer of silver bromide and x-ray wavelength, and between speeds of Type A and C film and wavelength.

The results agree with those obtained by Pelc (1945). His grain yield figures are however higher than those reported in this paper by a factor of 8, due to an erroneous silver mass determination (private communication from Pelc).

In order to compare this result with the theoretical expectation (§2) reference may be made to the Figures 9 and 10. Figure 9 shows the relationship between wavelength and the values of the energy absorbed per unit area in a 10 μ layer of silver bromide (as calculated from equation (2) §2). Figure 10 shows the calculated values of the number of quanta absorbed per unit area in a 10 μ layer of silver bromide (using equation (4) §2) as function of the wavelengths. The experimental relative speeds for films Types A and C are expressed in values of D/r , where D is the density and r the number of roentgen units, and were adjusted to fit the value of the energy absorbed at a wavelength of 0.3 Å. (Figure 9) and to fit the number of quanta absorbed at a wavelength of 0.3 Å. (Figure 10). The speed values refer only to measurements taken within the straight line part of the density–exposure curve of Type A and C films.

By reference to Figure 9 it is seen that the speed values for Type A film coincide with the values of the energy absorbed over the complete range of wavelengths.

Since the quantum energy rises with shorter wavelength, the number of grains rendered developable per absorbed quantum must evidently be proportional to the quantum energy absorbed at each wavelength. It will be noted however from the curve of Type C film (Figure 9) that the relative speed values of this film do not agree so closely with the values of the absorbed energy over the range of wavelengths considered.

In Figure 10 it is seen that the deviation of the speed values from the number of quanta absorbed increases with shorter wavelengths. This indicates that more grains are rendered developable per absorbed quantum with short than with long wavelengths.

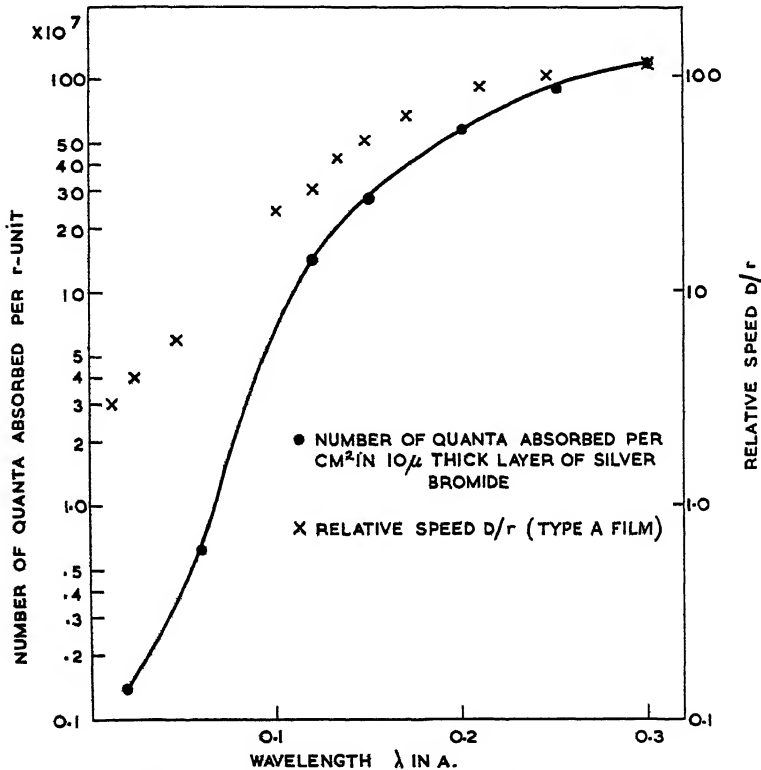


Figure 10. Relationships between the number of quanta absorbed per r-unit, relative speed of Type A film and x-ray wavelength.

The occurrence of the bend in the curves in Figures 6 and 7 at a wavelength of about 0.1 Å. is not understood. It may be a real effect, as it does not occur for Type C film. It is not likely to be an observational error since the same method and data were also used for the evaluation of Type C film (Figure 8).

The mechanism of the increase of the number of grains rendered developable with increasing quantum energy can be interpreted as follows: (a) With increasing quantum energy photoelectrons of increasing energy are released. The energy of the emerging electrons will be $h\nu - w_K$, where w_K is the energy required to remove an electron from the K level. (b) The range of photoelectrons thus released by a single quantum hit increases with quantum energy, thus affecting an increasing number of grains adjacent to the quantum absorbing grains.

The mean energies of photoelectrons plus recoil electrons released in a silver bromide emulsion for various wavelengths of incident radiations have been computed, and the corresponding ranges of electrons in gelatin and in silver bromide emulsions are known approximately. Experimental values of the latter were found in recent work on the range of electrons in nuclear track emulsions (Herz 1948, 1949b). At the critical wavelength of 0.37 μ (at a grain yield of unity), where the energy of photoelectrons appears to be insufficient to affect neighbouring grains, the mean electron energy has been found to be about 9 kv. A 9 kv. electron has a range of about 0.9 μ in a silver bromide emulsion of the Type A. It would thus lose its energy entirely before it reaches a neighbouring grain, as the spacings of the grains in Type A emulsion were found to be of the order of 1–2 microns. At a quantum efficiency of 2, corresponding to a wavelength of 0.23 μ for Type A film, the mean electron energy was found to be 32 kv., corresponding to a range of 9 μ in the silver bromide emulsion. This relatively large range appears evidently to be sufficient for the electrons to contribute to the developability of adjacent grains.

Summarizing, it may be said that about the same fraction of the energy is being utilized for latent image formation at all points in the range from low to high quantum energies. It is further interesting to note that as a consequence of this as many as 80 grains at 1,000 kv., and 100 grains in the range of gamma rays are rendered developable per one quantum absorbed.

(iii) *Rate of increase of grain yield with quantum energy.* The results of the grain yield of three different emulsions (Figure 5) are indicated by the different slopes of the curves. The rate of increase of grain yield with quantum energy is considerably steeper for Type C film than for Type A and B films, the slopes of which are fairly similar. It is believed that this higher rate of increase for Type C film is caused by the closer packing of grains in this emulsion. Considering the closer packing of grains in Type C film, it seems evident that for photoelectrons of the same range released from one quantum-absorbing grain, more grains may be affected than in an emulsion with looser packing of grains.

If a certain quantum energy (at about 0.05 μ) is reached, the slopes of the curves (Figures 6–8) become practically identical for the three emulsions. This probably means that if a certain length of path of photoelectrons is reached, the difference in spacing of the grains loses its importance in regard to grain yield, and similar yields are obtained for all three emulsions.

(iv) *Grain yield for type D film (Figure 5).* In Figure 5 it is shown that Type D film reaches a grain yield of unity only at relatively short wavelengths and is well below unity for wavelengths longer than 0.12 μ . This implies that the grains in Type D film are relatively insensitive. It appears that one quantum hit is not sufficient to render one grain developable, and that many grains, although affected, do not receive a sufficient number of ionizations to render them developable. Thus sub-images are formed, evidence of which was given by low intensity light post-exposure in § 6.

By comparing the results of quantum efficiency of Type D film in the region of grain yield below unity, at a wavelength of 0.12 μ , with those of films of Types A, B and C (Figure 5), it may be at first surprising that no flat region in the grain yield–frequency curve is obtained for this film. This may be understood by the following considerations. At wavelengths longer than 0.12 μ the range of the photoelectrons released is sufficient for them to reach grains adjacent to the

quantum-absorbing grain, as is shown by reference to the results obtained with other films. The energy of these photoelectrons is however insufficient to render the grains of Type D film developable (sub-image formation). At wavelengths shorter than 0.125 μ , it appears that not only the range but also the energy of the photoelectrons becomes great enough to produce sufficient ionizations and thus render neighbouring grains developable.

REFERENCES

- BURTON, P. C., and BERG, W. F., 1946, *Photogr. J.*, **86B**, 1-24.
 EGGERT, J., and NODDACK, W., 1927, *Z. Phys.*, **23**, 222.
 GÜNTHER, P., and TITTEL, H., 1933, *Z. Elektrochem.*, **39**, 646.
 HERZ, R. H., 1948, *Nature, Lond.*, **161**, 928; 1949 a, *Photogr. J.*, **898**, 89; 1949 b, *Phys. Rev.*, **75**, 478.
 MAYNEORD, W. V., 1940, *Brit. J. Radiolog.*, **13**, 235.
 NUTTING, P. G., 1913, *Phil. Mag.*, **26**, 423.
 PELC, S. R., 1945, *Proc. Phys. Soc.*, **57**, 523.
 SEEMANN, H. E., and MACGILLIVRAY, L. L., 1946, *Rev. Sci. Instrum.*, **17**, 12.
 SILBERSTEIN, L., and TRIVELLI, P. H., 1930, *Phil. Mag.*, **9**, 787.

The Analysis of Observations on Spaced Receivers of the Fading of Radio Signals

BY B. H. BRIGGS, G. J. PHILLIPS AND D. H. SHINN

Cavendish Laboratory, Cambridge

Communicated by J. A. Ratcliffe; MS. received 2nd August 1949

ABSTRACT. The fading of a radio wave once reflected from an irregular ionosphere is discussed in terms of the variable diffraction pattern produced at the ground. It is pointed out that fading may arise either by a drift of the diffraction pattern past a receiver, or by irregular variations in the pattern, or by both mechanisms together. It is shown how, from observations at three receiving points, it is possible to deduce the rate at which the pattern is changing and the velocity with which it drifts over the ground. This velocity is of interest as it may be related to the wind velocity at ionospheric heights in the atmosphere, but the relation of the diffraction pattern to the irregular ionosphere which produces it is not discussed in detail.

§ 1. INTRODUCTION

SEVERAL papers (Ratcliffe and Pawsey 1933, Pawsey 1935, and more recently Ratcliffe 1948 and Mitra 1949) have drawn attention to the fluctuations in the amplitude of a radio wave which has been singly reflected from the ionosphere. It has been pointed out that the reflection of a radio wave from an irregular ionosphere will lead to the production of an irregular 'diffraction pattern' on the ground. For practical reasons this pattern can only be observed at a few fixed points, by means of spaced receivers. If the ionosphere remained constant and fixed in position, the amplitude of the wave at each point on the ground would remain constant, and there would be no fading.

One way in which fading could be produced would be by a movement of the diffraction pattern past the observing points, as would happen if there were a horizontal wind in the ionosphere. In this case the fading recorded at two points

separated in the direction of the wind would be exactly similar but displaced in time, and from the time difference the velocity of the wind could be deduced.* If, now, the receivers were also gradually separated in the direction perpendicular to the direction of the wind, the recorded fading would become less similar, until finally no recognizable similarity would remain.

Another way in which fading could be produced would be by *random* changes in the ionosphere, as, for example, by the random motion of different parts of the ionosphere (Ratcliffe 1948). Such a mechanism would produce random changes of the diffraction pattern on the ground. The fading at two nearby receivers would then be similar, but would become more dissimilar as the receivers were separated from each other in any direction. There would be no tendency for the occurrence of similar fading with a time lag.

In general the two mechanisms mentioned above may operate simultaneously, so that, although the diffraction pattern moves on the ground with a steady drift velocity, it is subject to random changes of form as it moves. Under these circumstances the fading recorded by two receivers separated in the direction of the wind would be similar but displaced in time if the distance of separation were not too great, but the degree of similarity would decrease as the separation was increased, until at sufficient distance all similarity would have disappeared.

In this paper we shall show that, if certain plausible assumptions are made, it is possible to deduce the following from fading records taken at three points situated at the corners of a right-angled triangle: (a) the steady velocity with which the diffraction pattern drifts over the ground; (b) the rate at which the pattern alters as it moves; (c) the size of the irregularities in the pattern. The principle of the method is to consider the general similarity of the three records, and from the extent to which the differences exceed those expected from a drifting pattern of fixed character one can find the rate at which the pattern alters; hence one can make the correction that is necessary to the value of drift as it appears from the time lag. We need to consider the auto-correlation function of the fading records at the three points considered and also the cross-correlation functions between pairs of these records, and we devote the next section (§2) to a definition of these functions and a short discussion of them. The first step in solving our problem is to prove relations between certain fundamental quantities defined in §4, and this is done by two different methods which lead to the same results. Thus in §5 we discuss the correlations between the fading at the three receivers in a direct manner, focusing attention on the way in which the amplitude of signal varies with position and time, while in §6 we consider the form of a 'correlation surface' which shows how the correlation between the amplitudes at two receivers varies with their separation and with the difference in times of observation. It then remains to show in §7 how the results of the previous sections can be used to obtain the fundamental characteristics of the pattern from the three fading records.

The results which we deduce relate entirely to the diffraction pattern; in particular the steady drift and the random changes which we discuss are those of the pattern on the ground. We have deferred the interesting problem of relating this pattern to the phenomena in the ionosphere that give rise to it. By so doing we avoid having to base the discussion on any particular model of the

* If a simple diffraction mechanism is operative, the wind in the ionosphere will be half the velocity of drift of the diffraction pattern over the ground (Pawsey 1935).

ionosphere, although at certain stages in the analysis we do make some simplifying assumptions about the nature of the diffraction pattern. It is sufficient to remark that drifting effects in the diffraction pattern are probably connected with large scale movements in the ionosphere and changes in the pattern with random movements in the ionosphere.

Although we naturally consider the amplitude of a radio signal as the basic quantity when discussing fading, the results could be applied to any other suitably varying characteristic of an ionospheric reflection. Munro (1948) has shown how the magnitude of the magneto-ionic splitting, for example, is useful in the investigation of movements in the F region.

Some practical results concerning the drift of patterns over the ground have already been obtained (Mitra 1949). These experiments are being continued, with modified apparatus, and we hope to publish the results in the near future.

§ 2. MEASUREMENT OF CORRELATION

(i) *A Single Time-Record*

A record of fading, i.e. a graph of the amplitude R of a reflected wave against time t , is found to have a random character like the curve in Figure 1(a). An important function which arises in the study of a variable of this kind is the *auto-correlation function*. Thus, corresponding to the original random function $R(t)$, we have the auto-correlation function:

$$\rho(\tau) = \frac{\overline{\{R(t) - \bar{R}\}\{R(t+\tau) - \bar{R}\}}}{\overline{\{R(t) - \bar{R}\}^2}}. \quad \dots\dots (1)$$

This function measures the average correlation between values of R separated by a time interval τ . It has the value unity for $\tau=0$, and in all practical cases decreases smoothly at first as τ increases from zero. A plot of a typical auto-correlation function (or auto-correlogram) is shown in Figure 1(c). The value of τ at which $\rho(\tau)$ falls to zero is clearly that time separation at which two values of R have become completely independent.

In the analysis of fading of radio signals we need a suitable measure of the 'speed of fading'. It is possible to derive such a measure from the function ρ just defined, but it is more convenient to define the speed of fading, S , as follows:

$$S = \left| \frac{\partial \bar{R}}{\partial t} \right| / \bar{R}. \quad \dots\dots (2)$$

This is simply the mean rate of change of R , normalized to a constant mean value of R .

(ii) *Space Variations, and the Two-Dimensional Auto-Correlation Function*

By analogy with the definition of S we may measure the rate of variation of R over the ground at a given instant by a quantity G , where

$$G = \left| \frac{\partial \bar{R}}{\partial x} \right| / \bar{R}. \quad \dots\dots (3)$$

We shall assume that G is independent of the direction over the ground in which the variable x is measured. (An extension of the theory in which this condition is removed is mentioned in § 8.)

In the practical case the signal amplitude over the ground and in time will be represented by a function $R(x, y, t)$ of the two space coordinates x, y , and the time coordinate t . In order to simplify the analysis in the first instance, we shall introduce the artificial idea of a *one-dimensional* ground, so that R is a function of two variables only, x and t . We can derive the two-dimensional auto-correlation function of $R(x, t)$ by an obvious extension of equation (1). Thus

$$\rho(\xi, \tau) = \frac{\overline{\{R(x, t) - \bar{R}\}\{R(x + \xi, t + \tau) - \bar{R}\}}}{\overline{\{R(x, t) - \bar{R}\}^2}} \quad \dots\dots (4)$$

The example of a one-dimensional ground is useful for purposes of illustration because the functions $R(x, t)$ and $\rho(\xi, \tau)$ can be visualized as surfaces.

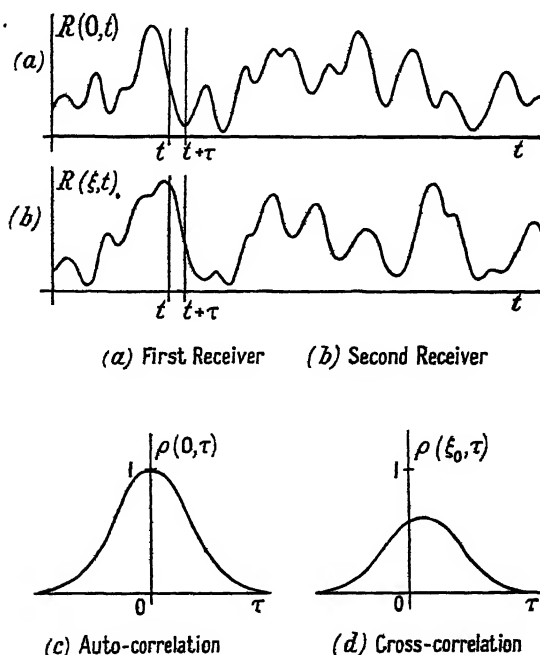


Figure 1.

In a hypothetical experiment on our one-dimensional ground we might obtain time records of the amplitude as shown in Figure 1 (a) and (b), at two fixed points, say at $x=0$ and $x=\xi_0$. We are then limited to a knowledge of $R(0, t)$ and $R(\xi_0, t)$. From either record the auto-correlogram corresponding to $\rho(0, \tau)$ and shown in Figure 1 (c) may be found. In addition, it is useful to evaluate the cross-correlation between the two records for different values of the time lag τ , so as to obtain the *lag-correlogram* corresponding to $\rho(\xi_0, \tau)$ shown in Figure 1 (d). These curves may be thought of as sections of the ρ surface by planes $\xi=0$ and $\xi=\xi_0$. Since in the practical case we know the amplitude variations in time much more completely than those along the ground, we must regard the averaging in

obtaining ρ from equation (4) as being carried out over time only. The physical problem is such that it may reasonably be assumed that the same result will be obtained whether we average over space and time, or over time only. This assumes that we possess statistically stationary records of sufficient length to include many cycles of the variation of R .

The extension of these ideas to cover a two-dimensional ground is straightforward. We then consider a correlation function $\rho(\xi, \eta, \tau)$ derived from $R(x, y, t)$ and $R(x + \xi, y + \eta, t + \tau)$.

§ 3. STATEMENT OF THE PROBLEM

Before proceeding with the mathematical discussion, and in order to introduce some of the difficulties encountered in the general case, we consider a two-dimensional diffraction pattern which is not subject to random changes. Consider the uniform drift of such a constant ground pattern past two observing points fairly close to each other. In Figure 2, let P and Q be two such points of observation

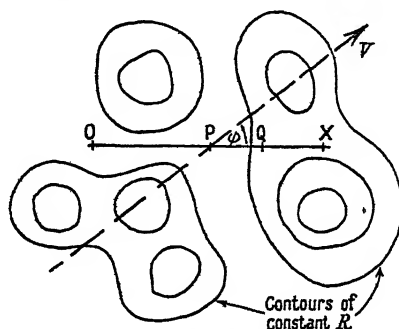


Figure 2.

on a line OX, and let the form of the drifting pattern be represented by contours of constant amplitude. We assume that, on the average, these contours are circles, i.e. the pattern changes equally rapidly in all directions. Let the pattern move in a direction making an angle ϕ with OX. Now the detail which passes P can never recur in exactly the same form at Q (unless ϕ is zero) but a somewhat similar variation is obtained at Q, with a time lag corresponding to the time taken for the pattern to traverse the projection of PQ on the line of motion of the pattern. As a result of the measurements at P and Q, then, we observe an *apparent* velocity V'_{OX} along OX. This quantity V'_{OX} is defined as the length PQ divided by the observed time lag, and is easily shown to be given by

$$V'_{OX} = V / \cos \phi. \quad \dots\dots (5)$$

It will be noted that this apparent velocity along OX is greater than the true velocity V of the pattern.

We shall find, however, that the velocity V_{OX} with which an observer would have to move along OX in order to reduce the speed of fading as observed by him to a minimum is given by

$$V_{OX} = V \cos \phi.$$

The difference between V'_{OX} and V_{OX} increases as ϕ increases.

The important conclusion to be drawn is that we have two methods of interpreting the observations along OX as a drift. The latter method is the one which

gives the true component of the total drift velocity along OX. But it may be more convenient in practice to make measurements leading to the apparent velocity V'_{OX} of the first method. If this is the case we must remember that we have found $V/\cos \phi$ and not $V \cos \phi$.

In the following sections we shall seek relations between the various velocities, like V'_{OX} and V_{OX} above, which arise in the more general case of a moving pattern with random changes, and consider how they are obtainable from records made on spaced receivers. It is possible to obtain the relations in more than one way. We shall therefore pass on immediately in the next section to the definitions of the velocities and other parameters which seem fundamental. Having introduced these, we shall in § 5 derive several useful formulae using a modification of the $R(x, t)$ surface which was mentioned in the discussion of the definition of $\rho(\xi, \tau)$. This approach may be called that of the $(x, V_c t)$ diagram. Then in § 6 a shorter alternative approach will be given which is somewhat more mathematical. Here the formulae are proved from the geometry of the $\rho(\xi, \tau)$ surface. In both these approaches we prove the results first for the 'one-dimensional ground' introduced in the previous section, and then generalize to the practical case of a two-dimensional ground.

We shall assume that the contours of constant ρ in the $\rho(\xi, \tau)$ diagram are similar ellipses concentric about the origin. The significance of this assumption is discussed later in § 8. Similarly when we consider a two-dimensional ground we shall take the constant ρ surfaces in the $\rho(\xi, \eta, \tau)$ diagram as similar concentric ellipsoids, which intersect the ξ - η plane in circles. In this way we include the assumption of an isotropic ground pattern.

§ 4. DEFINITIONS

The following definitions of various basic quantities are given in words so that they may apply generally in the cases of a one- or two-dimensional ground without being associated with any particular mathematical approach. In the illustration of the definitions, however, use is made of the correlation function for a one-dimensional ground, and it may be helpful to refer to Figure 4(c) which is a plan of the $\rho(\xi, \tau)$ surface showing the contour $\rho = \rho_0$, and to Figure 4(b) showing vertical sections corresponding to the correlograms of Figures 1(c) and 1(d).

(i) *Fading velocity* V'_c . This is a measure of the ratio of space shift to time shift needed to produce, on the average, the same change in the value of R . Thus

$$V'_c = x_0/t_0, \quad \dots\dots (6)$$

where x_0, t_0 satisfy $\rho(x_0, 0) = \rho(0, t_0)$.

Under the simplifying assumptions mentioned above this definition does not depend on the value of ρ chosen, or on the direction over the ground in which the space coordinate denoted by x is measured. Also, in view of (2) and (3), the definition leads to the expression

$$V'_c = S/G, \quad \text{or} \quad S = GV'_c. \quad \dots\dots (7)$$

We thus get the physical interpretation that the fading velocity V'_c is the velocity with which the unchanging ground pattern would have to move past an observer to give the observed speed of fading, i.e. the velocity of drift that would be needed to explain the fading *entirely* in terms of a drifting pattern with no random changes.

(ii) *Drift velocity* V . This is the velocity of an observer who has so adjusted his motion over the ground that he experiences the slowest possible speed of fading. Suppose he compares signal amplitudes at times τ_1 apart. His displacement ξ_1 during this time must be adjusted until $\rho(\xi_1, \tau_1)$ is a maximum, and then

$$V = \xi_1 / \tau_1. \quad \dots\dots (8)$$

This velocity is an unambiguous measure of the movement of the pattern along the ground. For a two-dimensional ground V has direction. Assuming axes of x and y on the ground let the direction be at angle ϕ to the x axis and let components of the velocity along the axes be V_x and V_y .

(iii) *Characteristic velocity*, V_c , and *characteristic speed of fading* S_c . These are the values of V'_c and S which are found by the observer above moving at velocity V , and give a quantitative measure of the amount of random change taking place. To this observer the ratio of space shift and time shift needed to produce a similar change in R is

$$V_c = x_0 / \tau_1, \quad \dots\dots (9)$$

where $\rho(x_0, 0) = \rho(V\tau_1, \tau_1) = \rho(\xi_1, \tau_1)$. This is seen by noting that any time shift τ must be accompanied by a displacement $V\tau$ for this observer. Also we have the relation corresponding to (7)

$$V_c = S_c / G, \quad \text{or} \quad S_c = GV_c. \quad \dots\dots (10)$$

(iv) *Apparent drift velocity*, V' , and *time lag* τ_0 for best cross-correlation. If the fading records of receivers separated by distance ξ_0 are examined for cross-correlation, i.e. if $\rho(\xi_0, \tau)$ is found for various values of τ , there will be a maximum correlation for a certain delay τ_0 . The apparent velocity of drift is then

$$V' = \xi_0 / \tau_0. \quad \dots\dots (11)$$

This velocity is analogous to the velocity V'_{0x} discussed in §3, and is the value that at first sight appears to be the true velocity of motion of the pattern, assuming this to be along the line of the receivers. For two-dimensional ground, we suppose that we have receivers spaced ξ_0 along Ox and spaced η_0 along Oy , and that τ_{0x} and τ_{0y} are the respective lags for best correlation. Then the apparent components are given by

$$V'_x = \xi_0 / \tau_{0x} \quad \text{and} \quad V'_y = \eta_0 / \tau_{0y}. \quad \dots\dots (12)$$

It is important to realize that these are not components of a vector V' , but two quantities like V' defined from observational data.

§ 5. DEDUCTIONS FROM THE (x, V_c, t) DIAGRAM

(i) *Construction of the Diagram*

We imagine plotted, by means of contours of R , the complete account of the happenings over a one-dimensional ground as time progresses. In general we obtain something like Figure 3(a). There is a tendency for the contours to be parallel to a line whose slope depends on the velocity of drift of the maxima and minima of R over the ground. If the origin of the x coordinate is moved along with velocity V we obtain Figure 3(b), corresponding to the state of affairs seen by the observer moving with velocity V , who experiences the minimum speed of fading. We have seen that for such an observer the ratio of the space and time shifts which have equal effect (on the average) in changing R is V_c . Therefore

we produce the third diagram in Figure 3 (c), using the coordinate $V_c t$ in place of t so that the average gradient of the contours is the same along both axes of the diagram.

Now we assume that in this diagram the average change, or expected correlation, between values of R at any two representative points depends only on their distance apart on the diagram. In other words a correlation surface corresponding to this new diagram would have contours in the form of circles centred on the origin. It can be seen by the reversal of the transformation back to Figure 3 (a), that on this assumption the real $\rho(\xi, \tau)$ contours must be ellipses.

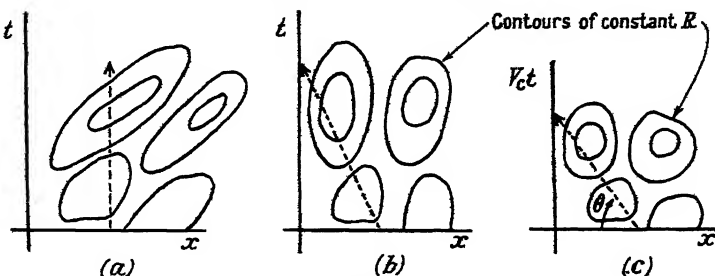


Figure 3.

Dotted lines cover the values of R experienced by a fixed point.

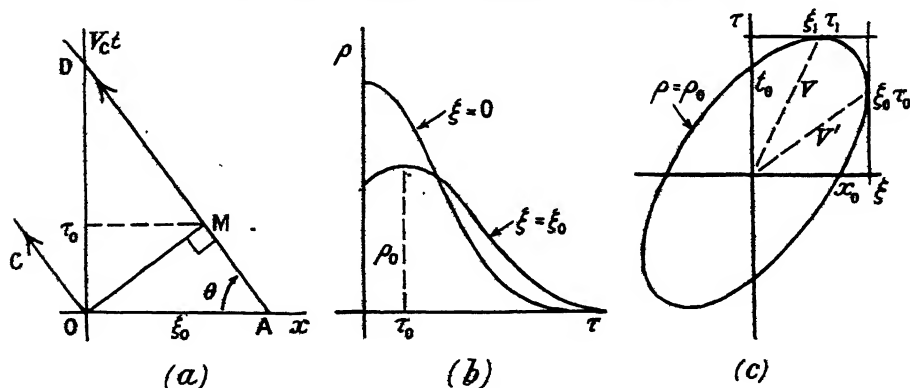


Figure 4.

A fixed point on the ground will have velocity $-V$ relative to the axes of Figure 3 (c). Its representative point on the diagram must move towards the negative x direction as time t increases; in fact it traverses a line sloping backwards at an angle θ where

$$\cot \theta = x/V_c t = V/V_c. \quad \dots (13)$$

Thus the values of R along this line are those experienced at the fixed point on the ground. The $(x, V_c t)$ diagram is redrawn in Figure 4 (a), in which the contours have been omitted for clarity.

(ii) Expression for the Speed of Fading

Since for a given time interval the distance along OC is greater by a factor $1/\sin \theta$ than along the $V_c t$ axis, we have a comparison of the corresponding speeds of fading S and S_c . Thus

$$S = S_c / \sin \theta = S_c [1 + (V/V_c)^2]^{1/2}$$

using (13) to eliminate θ . If we consider V_c' and V_c as measures of the speed of fading we have in view of (7) and (10)

$$V_c' = V_c[1 + (V/V_c)^2]^{\frac{1}{2}}.$$

The last two equations may be re-written

$$S^2 = S_c^2 + G^2 V^2, \quad \text{and} \quad V_c'^2 = V_c^2 + V^2. \quad \dots\dots (14)$$

The expression for the resultant speed of fading thus involves a square law combination of the contributions S_c and GV due respectively to random changes and steady drift.

(iii) Time Lag for Optimum Cross-Correlation

We obtain the fading record of a second receiver at distance ξ_0 from the first by following a sloping line AD in Figure 4(a) parallel to the first line OC and intercepting the x axis at A, a distance ξ_0 from O. It is seen that a value of R at O on the first record would be expected to be most nearly correlated with the closest point M on the other record. The height of M above the axis will give $V_c\tau_0$, where τ_0 is the time lag corresponding to optimum cross-correlation. From the geometry of Figure 4(a) we arrive at the equation

$$V_c\tau_0 = \xi_0 \sin \theta \cos \theta = \xi_0 V_c V / (V_c^2 + V^2)$$

in which we again make use of (13) to eliminate θ . Thus

$$V' = \xi_0 / \tau_0 = (V_c^2 + V^2) / V$$

or

$$V'V = V_c^2 + V^2,$$

and hence from (14)

$$V'V = V_c'^2. \quad \dots\dots (15)$$

It will be noticed that the determination of τ_0 gives V' immediately, but that V_c' is required to find V . This can be found as follows: the cross-correlation $\rho(\xi_0, 0)$ between the two records gives a measure of the effect of a pure space shift ξ_0 , and if we find the time lag τ_s which gives the same value of the correlation coefficient $\rho(0, \tau_s)$ measured on one record, then the definition of V_c' gives

$$V_c' = \xi_0 / \tau_s. \quad \dots\dots (16)$$

(iv) Other Methods

It is useful to consider the time τ_e for which the auto- and cross-correlograms cross one another when plotted on the same axes as in Figure 5(b). Mathematically τ_e may be defined as satisfying

$$\rho(0, \tau_e) = \rho(\xi_0, \tau_e) = \rho_e.$$

If we imagine the function $\rho(\xi, \tau)$ to be represented by a surface, then Figure 5(b) represents the section by planes $\xi = 0$ and $\xi = \xi_0$. The contour $\rho = \rho_e$ of the degree of correlation in this case is shown in Figure 5(c), which is a plan of the surface looking down on the (ξ, τ) plane.

The interpretation of τ_e in the $(x, V_c t)$ diagram of Figure 5(a) is that it corresponds to the point E on AD such that $OE = AE$, since with this value of time lag the auto-correlation corresponding to AE equals the cross-correlation corresponding to OE. By simple geometry D is twice the height of E above the x axis so that OD is $2V_c\tau_e$. Thus we have $\cot \theta = \xi_0 / 2V_c\tau_e$. But from (13) $\cot \theta = V/V_c$.

Hence

$$V = \xi_0 / 2\tau_e. \quad \dots\dots (17)$$

If the complete correlation curves can be worked out and plotted, the parameters V and V_e' are best obtained by the following method (Figures 6 (a) and (b)). Here a general point P on the record AD is considered with time lag τ relative to O or A . The auto-correlation for lag τ is thus governed by distance AP , and the cross-correlation for lag τ by distance OP . We have $AP = V_e \tau / \sin \theta$. Also let Q be a point on the line AD corresponding to time τ' , so that AQ gives the auto-correlation for lag τ' . Now make $AQ = OP$, i.e. let us define τ' as the time lag giving an auto-correlation equal to the cross-correlation for lag τ . Then we also have

$$OP = AQ = V_e \tau' / \sin \theta.$$

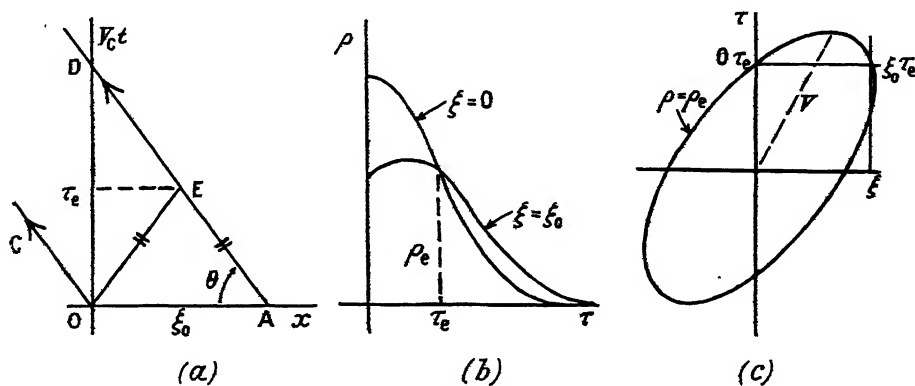


Figure 5.

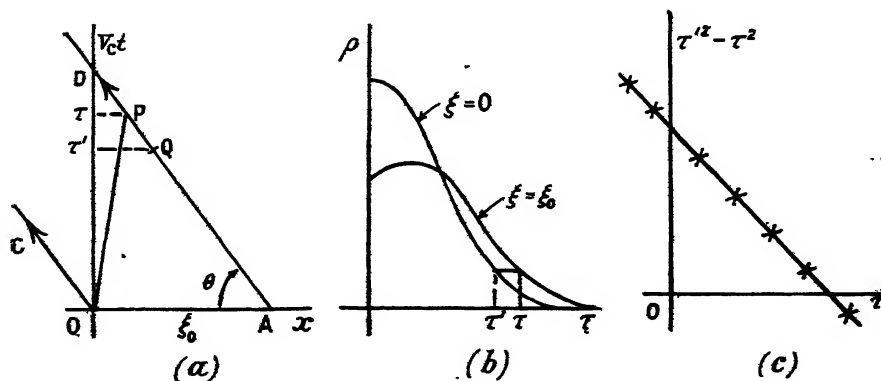


Figure 6.

From equations (13) and (14) $V_e / \sin \theta = V_e'$, and in triangle OPA of Figure 6 (a)

$$OP^2 = AP^2 + OA^2 - 2OA \cdot AP \cos \theta.$$

Combination of the last four results yields

$$V_e'^2 \tau'^2 = V_e'^2 \tau^2 + \xi_0^2 - 2\xi_0 \cos \theta \cdot V_e \tau / \sin \theta.$$

Hence, using (13),

$$\tau'^2 - \tau^2 = (\xi_0^2 - 2V\xi_0\tau) / V_e'^2. \quad \dots\dots (18)$$

This shows how use may be made of the complete auto- and cross-correlation diagrams to plot a straight line graph as shown in Figure 6(c), in which $\tau'^2 - \tau^2$ is plotted against τ . The intercepts which the line makes on the two axes are readily shown to be equal to $\xi_0/2V = \tau_e$ and $\xi_0^2/V_c'^2 = \tau_s^2$. Again V and V_c' may be obtained fairly simply from ξ_0 . The region where a good straight line is obtained is an indication of the region in which the assumptions of the present treatment are valid.

(v) *Extension for Two-Dimensional Ground*

Some of the above equations need modification before they will apply for a two-dimensional ground, unless the two receivers happen to be in line with the drift velocity. To obtain more general formulae we consider the natural extension of the $(x, V_c t)$ diagram to the $(x, y, V_c t)$ diagram shown in Figure 7. The fading

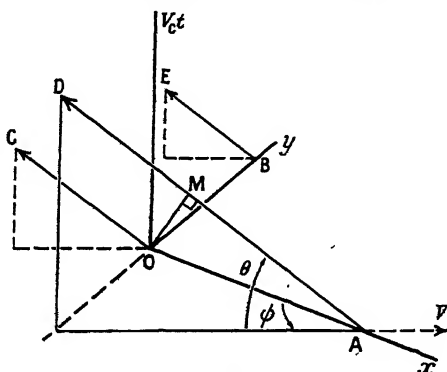


Figure 7.

records of three receivers at A, O and B are given by the intersection of the three sloping lines OC, AD and BE with the surfaces of constant R in $(x, y, V_c t)$ space. It is assumed that variations of R are equal on the average for equal displacements in any direction in this three-dimensional diagram. Considerations similar to those employed above lead to the results given below; the geometrical relation $\cos OAD = \cos \theta \cos \phi$ is required, the factor $\cos \phi$ converting V into its component V_x along Ox .

$$V_x V_x' = V_c'^2, \quad \dots\dots (19)$$

$$V_x = V \cos \phi = \xi_0/2\tau_{ex}, \quad \dots\dots (20)$$

$$V_x' = V'/\cos \phi = \xi_0/\tau_{0x}, \quad \dots\dots (21)$$

$$\tau'^2 - \tau^2 = (\xi_0^2 - 2V_x \xi_0 \tau)/V_c'^2. \quad \dots\dots (22)$$

In addition (14) and (15) still apply, V and V' being quantities that would be obtained directly by applying the formulae for a one-dimensional ground to records of two receivers in line with the drift V . τ_{ex} and τ_{0x} are written for the values of τ_e and τ_0 deduced from the records of the receivers on the x axis. A similar set of equations holds for the y axis, with the appropriate value η_0 for the receiver spacing in place of ξ_0 , and with $\sin \phi$ replacing $\cos \phi$. Also we have

$$\tan \phi = V_y/V_x = V_y'/V_x' \quad \dots\dots (23)$$

which, if $\eta_0 = \xi_0$, leads to the simple result

$$\tan \phi = \tau_{ex}/\tau_{ey} = \tau_{0y}/\tau_{0x}.$$

It is interesting to note that the equations for direction of V are not complicated by random changes; they are just the equations that would be expected in the special case considered in § 3.

We conclude from the results of this section that fading records taken at three receivers placed at corners of a right-angled triangle will be sufficient to enable us to solve our problem completely.

§ 6. GEOMETRY OF THE ρ -SURFACE

(i) *One-Dimensional Ground; Relations arising from Optimum Cross-correlation*

In this section we shall show how the results of equations (14) and (15) may be arrived at differently by a consideration of the geometry of the space-time correlation surface $\rho(\xi, \tau)$. As previously mentioned contours on the ρ -surface will be taken as ellipses of the form shown in Figure 4(c).

Suppose two fading records are made at two points separated by a distance ξ_0 , and it is required to find the time lag τ_0 which would produce the best cross-correlation between them. The problem can be solved by drawing a line on Figure 4(c) through the point $(\xi_0, 0)$ parallel to the τ -axis, and determining the value of τ_0 where it touches one of the elliptical contours ρ_0 . This value ρ_0 will then be the maximum value of the cross-correlation; let t_0 be the value of τ where this ellipse cuts the τ axis, so that t_0 is the time lag for *auto-correlation* required to give a value of correlation equal to this maximum *cross-correlation* corresponding to a lag τ_0 .

x_0 is the intercept on the ξ axis, and (ξ_1, τ_1) is the point at which the tangent parallel to the ξ axis meets the ellipse. This notation accords with the symbols used in the definitions of § 4 so that equations (6), (8), (9) and (11) may be restated as our mathematical definitions:

$$V'_e = x_0/t_0, \quad V = \xi_1/\tau_1, \quad V_e = x_0\tau_1, \quad V' = \xi_0/\tau_0. \quad \dots\dots (24)$$

Let the ellipse be $A\xi^2 + B\tau^2 + 2H\xi\tau = 1$.

If it has a vertical tangent touching at (ξ_0, τ_0) and passes through $(0, t_0)$, then

$$A = \frac{t_0^2 + \tau_0^2}{t_0^2 \xi_0^2}, \quad B = \frac{1}{t_0^2}, \quad H = \frac{-\tau_0}{t_0^2 \xi_0}. \quad \dots\dots (25)$$

And from the fact that it meets a horizontal tangent at (ξ_1, τ_1) and also passes through $(x_0, 0)$, we may deduce

$$\tau_1^2 = \frac{A}{AB - H^2}, \quad \xi_1 = -\frac{H\tau_1}{A}, \quad x_0^2 = \frac{1}{A}. \quad \dots\dots (26)$$

Equations (24), (25) and (26) enable us to express the various velocities in terms of A , B and H or in terms of our observables ξ_0 , τ_0 and t_0 :

$$\left. \begin{aligned} V'_e &= \left(\frac{B}{A}\right)^{\frac{1}{2}} = \frac{\xi_0}{(t_0^2 + \tau_0^2)^{\frac{1}{2}}}, & V &= -\frac{H}{A} = \frac{\xi_0\tau_0}{t_0^2 + \tau_0^2}, \\ V_e &= \frac{[AB - H^2]^{\frac{1}{2}}}{A} = \frac{\xi_0\tau_0}{t_0^2 + \tau_0^2}, & V' &= -\frac{B}{H} = \frac{\xi_0}{\tau_0}. \end{aligned} \right\} \quad \dots\dots (27)$$

From (27) it is readily shown that

$$VV' = V_e'^2 = V_e^2 + V^2 \quad \dots\dots (28)$$

which is equivalent to (14) and (15). In addition we have an alternative method of finding V'_e , and hence V and V_e , because we have taken t_0 as observable;

comparison of the expression for V_e' in (27) with (16) shows that the connection with the previous method is given by

$$\tau_s^2 = t_0^2 + \tau_0^2. \quad \dots\dots (29)$$

This relation may be demonstrated more directly by showing that the times τ_s , t_0 and τ_0 are proportional to the sides OA, OM and AM of the right-angled triangle OMA in the $(x, V_e t)$ diagram of Figure 4(a).

(ii) Other Methods

The time τ_e is defined so that $\rho(0, \tau_e) = \rho(\xi_0, \tau_e)$. The ellipse $\rho = \rho_e$ that passes through points $(0, \tau_e)$ and (ξ_0, τ_e) is shown in Figure 5(c). The mid-point of the horizontal chord between these points must lie on the same radial as the point of contact of the horizontal tangent. But we have seen that such a radial has slope $1/V$, so that

$$V = \xi_0 / 2\tau_e. \quad \dots\dots (17)$$

We can also prove (18) from the geometry of an ellipse; if $A'\xi^2 + B'\tau^2 + 2H'\xi\tau = 1$ is the ellipse passing through (ξ_0, τ) and $(0, \tau')$, then $B'\tau'^2 = A'\xi_0^2 + B'\tau^2 + 2H'\xi_0\tau$ or $\tau'^2 - \tau^2 = (A'\xi_0^2 + 2H'\xi_0\tau)/B'$. Now since all correlation ellipses have been assumed similar, we may use ratios A/B and $2H/B$ for the ellipse $\rho = \rho_0$ discussed in (i) above. From (27) $A/B = 1/V_e'^2$ and $2H/B = -2/V' = -2V/V_e'^2$, and using these values in the expression above we get (18).

(iii) Extension for Two-Dimensional Ground

As R must now be regarded as a function of three variables, x, y, t , we have to consider the corresponding correlation function

$$\rho(\xi, \eta, \tau) = \frac{\{R(x + \xi, y + \eta, t + \tau) - \bar{R}\}\{R(x, y, t) - \bar{R}\}}{\{R(x, y, t) - \bar{R}\}^2}. \quad \dots\dots (30)$$

The equivalent of our previous assumption of ellipses for lines of constant $\rho(\xi, \tau)$ is the assumption that surfaces of constant $\rho(\xi, \eta, \tau)$ are concentric ellipsoids centred on the origin of the coordinates. We also assume that the ground pattern is statistically isotropic, which implies that each correlation ellipsoid intersects planes perpendicular to the τ axis in circles.

Now we can consider in the same way as before the elliptical sections in the vertical planes $\eta = 0$, $\xi = 0$, and the vertical plane containing the direction of drift. Analysis of the latter section leads to (28), and to expressions for the velocities in terms of observations at points in line with the drift. Since this would not in general be of immediate use we must consider the other sections.

Figure 8 shows the vertical sections of the ellipsoid $\rho = \rho_{0x}$ corresponding to the maximum correlation between records of the two receivers on the x -axis. The $\eta = 0$ section therefore meets a vertical tangent at $(\xi_0, 0, \tau_{0x})$ and we have $V_x' = \xi_0/\tau_{0x}$ as in (12); also, following the previous treatment, x_0 and t_{0x} are the intercepts on the ξ and τ axes, and $(\xi_1, 0, \tau_{1x})$ is the point of contact of the horizontal tangent.

The point (ξ_2, η_2, τ_2) where a horizontal plane is tangential to the ellipsoid gives

$$\tan \phi = \eta_2/\xi_2, \quad V = (\xi_2^2 + \eta_2^2)^{1/2}/\tau_2,$$

i.e. the components are $V_x = \xi_2/\tau_2$, $V_y = \eta_2/\tau_2$ as may be seen from the definition of V . Because the ellipsoid has circular horizontal sections it may be shown that V_x is also given by $V_x = \xi_1/\tau_{1x}$ which means that in this case

V_x can equally well be defined as the movement *along the x axis* that gives a minimum speed of fading.

Similar considerations apply to the receivers on the y axis which show a maximum cross-correlation $\rho = \rho_{0y}$ when the time lag is τ_{0y} and we obtain an elliptical section in the $(\eta\tau)$ plane with vertical and horizontal tangents at $(0, \eta_0, \tau_{0y})$ and $(0, \eta_1, \tau_{1y})$ and with intercepts y_0, t_{0y} . However, since all sections in this plane will be similar figures, we have scaled down the section in Figure 8 by a factor $k = x_0/y_0$ so that the diagram may represent for simplicity the sections of one ellipsoid having a circular cross section of radius x_0 in the $(\xi\eta)$ plane.

The remaining relations $V'_c = x_0/t_{0x}$, $V_c = x_0/\tau_2$ are straightforward statements of the definitions.

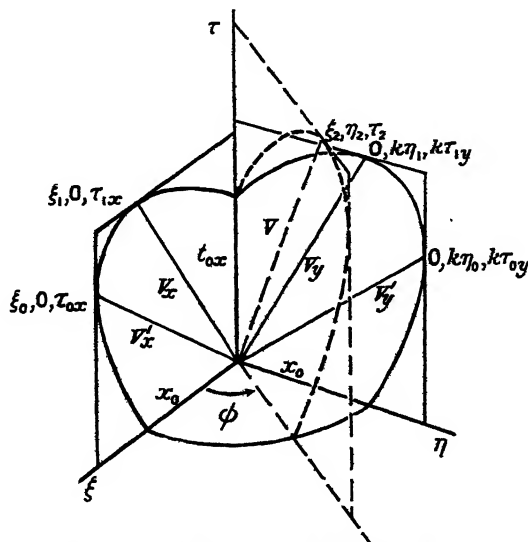


Figure 8. Sections of the ellipsoid $\rho = \rho_{0x}$

It is only necessary to consider the coordinate geometry of the ellipsoid in a way similar to that used for the ellipse in the case of one-dimensional ground to demonstrate from the mathematical definitions the results (19) to (23) of § 5. Also in (27) the expressions for V and V' in terms of t_0 and τ_0 are valid with suffixes x after these four symbols; the expression for V'_c is similarly valid, but the equation for V_c must be changed so that it is equivalent to

$$V_c^2 = V_c'^2 - V_x^2 - V_y^2, \quad \dots\dots (31)$$

a relation which follows from (14) and the fact that V_x, V_y are components of V .

§ 7. DISCUSSION OF RESULTS AND ANALYSIS OF RECORDS

We now turn to the way in which the results of the previous sections can be applied to the analysis of a set of three fading records. We are usually most interested in a determination of the velocity of drift V of the diffraction pattern since this provides evidence of steady motion somewhere in the ionosphere. But we may also be interested in the values of V_c as a measure of random changes, and in the factor which defines the fineness of structure of the diffraction pattern because these quantities have obvious importance in the study of fading.

The method of calculating V will depend on whether facilities or time permit a computation of auto- and cross-correlations of the records for a sufficient

number of values of τ . Sometimes the time lags, τ_{0x} and τ_{0y} , for the best cross-correlation between records of receivers on the x and y axes respectively can be estimated fairly accurately by eye, without having to find ρ . If this is the case it is simplest to find the direction of V straight away from the equation $\tan \phi = \xi_0 \tau_{0y} / \eta_0 \tau_{0x}$ and to calculate V' from $1/V'^2 = \tau_{0x}^2 / \xi_0^2 + \tau_{0y}^2 / \eta_0^2$. These equations follow from (23), (21) and the corresponding expression for V_y' . We then need to obtain V from V' . If we recall the relation between the velocities $VV' = V_c'^2$ (equation (15)) we see that it is sufficient to find V_c' . As this quantity is the ratio of the speed of fading S to the factor G of the fineness of the ground pattern, it may be found from the auto-correlogram and the value of cross-correlation with no time shift (see equation (16)). Should we find V_c' approximately equal to V' we notice that either of these two quantities gives the approximate magnitude of V ; this is quite a likely occurrence if, as has been supposed, the time lags of the records can be estimated visually.

If we perform a correlation analysis of the records we have the possibility of finding the time shift for optimum cross-correlation and, proceeding as in the previous paragraph, first finding V' and V_c' and then deducing the value of V . However, there are two alternative methods which give the components of V directly. The first method requires several values of both the auto- and cross-correlation so that the shift τ_c for which the correlations are equal may be found (see equation (20)). As a second alternative, a rather more thorough analysis will enable the graphical method described in §5(iv) to be used, and then the best values of V_c' as well as of the components of V will be obtained (see equation (22)). The value of the factor G is obtainable from V_c' and the speed of fading S by equation (7), $V_c' = S/G$.

In order to obtain the characteristic velocity V_c (which is a measure of the random changes) we must first find V_c' and the magnitude and direction of V , or V_x and V_y , and use equation (14) or (31). If V is large compared with V_c , so that V_c' becomes practically equal to V we may say that the drift is the predominant cause of fading. On the other hand if V_c is larger than V the fading is mainly due to random changes, and V becomes more difficult to find.

It is important in these calculations to perform some checks on the basis of the theory by comparing the two values for V_c' obtainable from the two pairs of stations. For if our assumption of an isotropic ground pattern holds the values should be equal. We should also obtain a third value of V_c' by exactly similar analysis of the third pair of records given by receivers A and B, and only if all three values agree can we be sure of isotropy. If the values are not in agreement a modified method of calculation is required which will be discussed in §8.

One further point in these calculations is that it is not essential to evaluate the correlation function fully in order to find a maximum or equality of correlation; in practice it may be found that other methods of measuring correlation are simpler.

§8. LIMITATIONS AND EXTENSION OF THE THEORY

In order to discuss the nature of the assumptions involved in the analysis, we return to a further consideration of the correlation function $\rho(\xi, \tau)$ for a one-dimensional ground (Figure 4(c)). Contours of constant ρ have been assumed to be ellipses. There is no obvious physical reason why this should be so, and

it is conceivable that the contours might be of any shape whatever. However, the ρ -surface certainly has a maximum at the origin, where the correlation is necessarily unity. Now it is a theorem in analysis that sufficiently near a maximum or minimum, any surface has elliptical contours provided the principal radii of curvature are not infinite or zero. Consequently our assumptions will normally be valid in the limit for sufficiently small values of ξ and τ .

In considering the problem for a two-dimensional ground, we have assumed up to the present that the diffraction pattern was statistically isotropic so that at a given instant the auto-correlation function $\rho(\xi)$ as a function of distance is the same in all directions over the ground. This assumption was only made in order to simplify the analysis, and there is no difficulty in principle in extending the results to cover the possibility of a non-isotropic ground pattern. From the preceding paragraph it will be seen that the analysis will be made one stage more general if we allow the contours of equal correlation on the ground to have the form of ellipses. This defines two principal directions on the ground which we may conveniently take as axes of x and y . These axes will not, of course, have any relationship to the lines along which the observing receivers are placed. We now have two principal values of G , measured along the two axes, which we denote by G_x and G_y , and two values of V_c' , defined by $V_{cx}' = S/G_x$ and $V_{cy}' = S/G_y$. The quantities S , S_c and V remain uniquely defined and require no generalization. An extension of the previous analysis enables all the above quantities to be found from fading records taken with three receivers placed at the corners of a right-angled triangle of arbitrary orientation. Since this extension involves no new principles it will not be given here.

A non-isotropic ground pattern of the type discussed above will be expected to occur whenever some direction in the ionosphere is specially distinguished. For example, in reflection at oblique incidence, the plane of propagation defines a special direction, and even at vertical incidence a special direction may be provided by the direction of the drift, or by the direction of the Earth's magnetic field.

Finally, we should point out that our analysis is only applicable on occasions when statistically stationary records of sufficient length are obtainable.

ACKNOWLEDGMENTS

The work here described forms part of a programme of Radio Research at the Cavendish Laboratory, supported by a grant from the Department of Scientific and Industrial Research.

We wish to thank our colleague Mr. S. A. Bowhill for helpful suggestions, particularly in connection with the methods of analysing records. We are also indebted to the Department of Scientific and Industrial Research for the individual maintenance allowances awarded to each of us.

REFERENCES

- MITRA, S. N., 1949, *Proc. Instn. Elect. Engrs.*, Pt. III, **96**, 441.
 MUNRO, G. H., 1948, *Nature, Lond.*, **162**, 886.
 PAWSEY, J. L., 1935, *Proc. Camb. Phil. Soc.*, **31**, 125.
 RATCLIFFE, J. A., 1948, *Nature, Lond.*, **162**, 9.
 RATCLIFFE, J. A., and PAWSEY, J. L., 1933, *Proc. Camb. Phil. Soc.*, **29**, 301.

Les renforcements brusques des ondes très longues

PAR R. BUREAU

Laboratoire National de Radioélectricité, Bagneux, France

Communicated by J. A. Ratcliffe; MS. received 26th July 1949. Paper read at Physical Society Conference at Cambridge, 14th-16th July 1949

ABSTRACT. During the last twenty years, enhancements of the mean level of atmospherics, on 11 km. wavelength, have been associated with some other phenomena such as simultaneous fade-outs of the field intensity of stations on decametric wavelengths, perturbations of terrestrial magnetism or Wolf numbers relative to solar activity.

More recently a search has been made for occasions on which the association between the different phenomena is not so close. It has been found that the usual fade-out on short waves does not occur in the same manner on all stations recorded; sometimes, indeed, an enhancement of short duration takes place at the beginning of the expected fade-out. This enhancement is probably due to the reception of a radioelectric radiation from the sun.

The author gives some examples of such phenomena which can be used for a better forecasting of radioelectric disturbance due to abnormal solar activity.

LES renforcements brusques des ondes très longues (kilométriques et myriamétriques) et tout particulièrement ceux du niveau des atmosphériques ont joué et jouent encore un rôle essentiel dans l'étude des perturbations ionosphériques à début brusque (P.I.D.B.).

Cette étude a débuté par la recherche des synchronismes. Ainsi a-t-on groupé autour des P.I.D.B. des manifestations très diverses; ce travail fut facilité par la brusquerie avec laquelle s'installent les phénomènes. Donnons-en deux exemples parmi beaucoup d'autres :

le 21 février 1942 (figure 1*) début brusque à 13 h. 26 sur le niveau moyen des atmosphériques et, simultanément, crochet sur la déclinaison du magnétisme terrestre (exemple fameux, synchrone avec une éruption chromosphérique intense se détachant sur le bord E du soleil);

le 28 mai 1936 à 14h. et à 18h. T.U., (figure 2) disparition des échos ionosphériques à Huancayo et en synchronisme deux renforcements des atmosphériques à Rabat.

Autre mode de comparaison, mais à une autre échelle : établir la liste interrompue des P.I.D.B. Ce travail s'accommode très bien de l'enregistrement des atmosphériques sur 11,000 mètres. On met ainsi en évidence l'influence du cycle solaire undécennal; la figure 3 donne le nombre annuel des renforcements pendant 20 ans; on y retrouve deux maximums successifs correspondant aux maximums d'activité solaire.

On entre maintenant dans une nouvelle période de la recherche; celle qui se base sur l'étude des divergences. Ici encore deux méthodes s'offrent à nous :

- (a) une méthode statistique où l'on relève le pourcentage d'apparitions synchrones des diverses manifestations rattachables aux P.I.D.B.
- (b) l'étude d'exemples individuels.

* For all photographs see group of Plates following p. 128.

Ce sont des exemples de cette dernière méthode que nous donnerons ci-après en choisissant quatre cas qui se rattachent aux rayonnements radioélectriques solaires. Il était indiqué de rechercher si certains de ces phénomènes nouveaux, dont l'apparence rappelait si étrangement les P.I.D.B., n'étaient pas en liaison avec elles, tout au moins partiellement. C'est ce qu'ont fait Appleton et Hey qui ont montré qu'il y avait une concordance remarquable entre les deux phénomènes pendant une période d'activité solaire intense qui s'est déroulée du 30 janvier au 14 février 1946. Or, comme il a été signalé par Appleton (1945, 1946), il peut

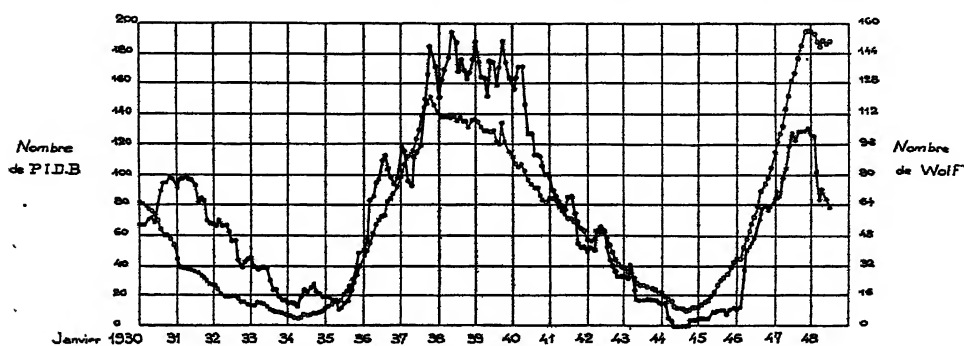


Figure 3.

—○—○—○—○—○— Moyennes mobiles annuelles de l'activité solaire (nombre de Wolf).
 —●—●—●—●—●— Moyennes mobiles annuelles du nombre des P.I.D.B.

Remarque. Les moyennes concernent les 12 mois précédant chaque point.

y avoir un antagonisme dans les actions sur l'ionosphère, provoquées au même moment, les unes par les rayonnements ultra-violet, les autres par le rayonnement électromagnétique sur ondes métriques et décamétriques. Je verse au dossier de cette enquête quatre documents dans lesquels semble bien se manifester cet antagonisme. Le premier, que j'ai déjà signalé (Bureau 1947), relatif au 8 février 1946 et trois autres relatifs au 28 décembre 1948, 5 avril 1949 et 7 mai 1949.

8 février 1946 (figure 4)

La courbe supérieure est celle du niveau moyen des atmosphériques sur 11,000 mètres de longueur d'onde le 8 février 1946 de 09 h. à 13 h. T.U. La courbe inférieure donne le champ de Genève sur 48,66 mètres. On trouve là un exemple particulièrement net du renforcement sur 11,000 mètres accompagné d'un affaiblissement sur 48,66 mètres. Mais ce qui rend cet enregistrement particulièrement précieux c'est que l'évanouissement de Genève est précédé d'un renforcement brutal d'une durée au plus égale à quelques secondes (la plume n'a pas eu le temps d'inscrire un trait continu dans la montée). Il faut sans doute voir là l'exemple d'un sifflement que certains auteurs ont signalé parfois au début d'un évanouissement brusque. Le mécanisme tel que l'envisageait Appleton serait le suivant : au début de la perturbation le rayonnement radioélectrique solaire atteint la terre, après la traversée des régions E et F de l'ionosphère avant que le rayonnement ultra-violet n'ait tendu un écran ionisé dans la région D. Très peu de temps après, l'écran est tendu et l'évanouissement se produit. Le raisonnement d'Appleton s'applique à des fréquences comprises entre 10 et 30 Mc/s. (ondes métriques et décamétriques).

28 décembre 1948 (figure 5)

La courbe supérieure reproduit l'enregistrement des atmosphériques à Bagneux sur 11,000 mètres de longueur d'onde le 28 décembre 1948 de 09 h. à 16 h. T.U. La figure inférieure a trait à l'enregistrement à Bagneux du champ de l'émetteur de Leipzig sur 30,80 mètres. Le renforcement des atmosphériques n'est pas accompagné d'un évanouissement sur ondes décamétriques, mais il accuse une montée rapide, synchrone avec celle des atmosphériques sur 11,000 mètres. Cette montée cesse très peu de temps après et la plume redescend au niveau précédent. Aucun évanouissement à début brusque n'est signalé en Europe ni en Amérique le 28 décembre. Deux appareils totalement distincts ont enregistré à Bagneux la montée sur 11,000 mètres*. Se trouve-t-on là en présence du même phénomène que ci-dessus? On ne peut l'affirmer d'une manière absolue mais il était bon de faire connaître cet exemple.

5 avril 1949 (figure 6)

La figure 6 a trait à une partie de la journée du 5 avril 1949, à cheval sur le jour et la nuit. Le soleil s'est couché à 18 h. 26 à Paris et à 18 h. 51 à Rabat. A la fin de la journée ont apparu, séparées d'à peu près une heure, deux P.I.D.B. particulièrement importantes sur 11,000 mètres. La première colonne de la figure reproduit trois enregistrements distincts d'atmosphériques sur environ 11,000 mètres à Bagneux, ainsi qu'un quatrième relevé à Rabat. La seconde colonne reproduit pour la même période trois enregistrements d'ondes décimétriques effectués à Bagneux et correspondant aux émissions de (a) Leipzig sur 9,732 kc/s. (30,80 mètres), (b) et (c) Washington WWV sur 2,000 kc/s. (15 mètres). (La première courbe de Washington correspond au champ de l'onde porteuse, la seconde reproduit l'onde porteuse plus la modulation.) En bas de la seconde colonne, on trouve un extrait du goniogramme à secteur étroit de Trappes (environ 25 km. à l'ouest de Bagneux).

La première P.I.D.B. est du type habituel; début à 15 h. 18, maximum à 15 h. 29, fin vers 16 h. 05. Elle est tout aussi nette à Rabat qu'à Bagneux, après une période de brouillages qui s'y termine, heureusement, un peu avant 15 h. Le goniogramme de Trappes y est très sensible, surtout pour la direction du foyer principal, c'est à dire vers le Sud. Le champ de Leipzig semble complètement indifférent à la P.I.D.B. Le champ de Washington accuse un évanouissement dont le début semble bien concorder avec le départ du renforcement sur 11,000 mètres. La descente se termine à 15 h. 27. Des évanouissements ont été constatés sur le trafic commercial en particulier à la station de Radio-France à Villecresnes où l'on a signalé à 15 h. 15 un important évanouissement sur le trafic de New-York. Le C.R.P.L. signale un certain nombre d'affaiblissements mais indique comme point de départ 15 h. 22.

Cette première P.I.D.B. respecte les caractères généralement admis relatifs à l'effet inverse des ondes décimétriques et des ondes kilométriques. Tout au plus peut-on signaler que l'évanouissement de Washington à Bagneux est précédé d'une pointe de très courte durée particulièrement visible sur l'enregistrement mixte (onde porteuse et modulation). Ce renforcement fait songer à celui du 8 février 1946 décrit plus haut. Notons aussi que l'évanouissement signalé par le C.R.P.L. est de 4 minutes plus tardif que les heures de départ précitées.

* Un observateur se trouvait présent devant les appareils au moment de ce phénomène et a pu constater le synchronisme relatif ci-dessus.

La deuxième P.I.D.B. est apparue brutalement à 16 h. 25. Elle se compose vraisemblablement de deux P.I.D.B. superposées, mais le partage entre l'une et l'autre est simplement indiqué par un léger palier des courbes des atmosphériques enregistrées à Bagneux. La séparation paraît plus nette à Rabat. Les heures de départ sont à 16 h. 25 et 16 h. 35. La descente pour les deux P.I.D.B. commence à 16 h. 54 pour se terminer vers 17 h. 30 dans le minimum qui précède le coucher du soleil, minimum très marqué sur le goniogramme. Le départ de la première partie de la deuxième P.I.D.B. est marqué sur le goniogramme, à la fois sur le foyer Sud et sur le foyer Ouest-Nord-Ouest.

Voici maintenant ce que l'on a constaté sur les évanouissements. Sur Leipzig s'observe une montée continue oscillant autour d'une courbe médiane qui donne l'apparence d'une montée logarithmique et qui correspond sans doute à la diminution de l'absorption. Insensible, comme nous l'avons dit plus haut à la P.I.D.B. de 15 h. 18, le champ passe par un maximum entre 16 h. 32 et 16 h. 35. Il accuse ensuite une chute rapide qui se termine, avant d'avoir atteint un évanouissement total, et cela légèrement après le renforcement brusque sur 11,000 mètres. Sur Washington la courbe de Bagneux accuse une descente en deux parties avec un palier correspondant à celui inscrit sur 11,000 mètres. La chute de Leipzig commence au moment où finit ce palier. Des évanouissements sont signalés à 16 h. 38 à New-York et 16 h. 22 à Washington.

Après le coucher du soleil le goniogramme d'atmosphériques signale une montée nocturne qui concerne tout d'abord une puissante source aux environs du Sud. Mais comme on le sait la nuit il n'y a pas de P.I.D.B.

Dans cet exemple nous trouvons des anomalies seulement sur le champ de Leipzig qui n'éprouve pas d'évanouissement dans la première P.I.D.B. et qui, de plus, accuse un renforcement provisoire entre la première et la seconde partie de la deuxième P.I.D.B. Le champ de Leipzig s'affaiblit seulement au moment où apparaît la seconde partie de la dernière P.I.D.B.

7 mai 1949 (figure 7)

La courbe supérieure représente l'enregistrement des atmosphériques sur 11,000 mètres à Bagneux de 09 h. à 16 h. T.U. le 7 mai 1949. La courbe centrale donne l'enregistrement à Bagneux du champ de Washington WWV sur 15 mètres, la courbe du bas l'enregistrement de Moscou sur 26,60 mètres. Deux P.I.D.B. se manifestent à 10 h. 45 et 13 h. 45 T.U. Sur la première on retrouve un très rapide et bref renforcement des ondes décamétriques de WWV précédant l'évanouissement habituel. Un observateur se trouvait présent devant les enregistreurs et a constaté la chronologie des diverses phases du phénomène, et en particulier la simultanéité de la montée de WWV avec la montée des atmosphériques sur 11,000 mètres. La seconde P.I.D.B. est du type normal, l'évanouissement des ondes courtes concordant avec le renforcement des ondes longues.

Ces quelques exemples ont seulement pour but de mettre en lumière un phénomène qui peut être très important à l'avenir surtout dans les prévisions d'orages magnétiques ou d'orages ionosphériques. On sait en effet que dans un certain nombre de cas on peut associer une P.I.D.B. (due à un rayonnement ondulatoire solaire) au déclenchement ultérieur d'un orage magnétique (dû à un rayonnement corpusculaire), ce qui entraîne des troubles très profonds dans les

radiocommunications, le délai qui sépare les deux phénomènes étant en général de 36 heures et pouvant varier d'environ 20 heures à environ 40 heures. On conçoit ainsi à quel point il est intéressant de rattacher ces phénomènes du rayonnement à une troisième sorte de rayonnement dû au soleil et qui est un rayonnement radioélectrique décelable sur les ondes métriques, décimétriques et même centimétriques. Il est donc assez curieux de signaler que les P.I.D.B. si spectaculaires du 5 avril ont été signalées aux services radioélectriques comme annonçant des perturbations probables de l'ionosphère dans un délai allant de 36 heures à 50 heures. Or, le 8 avril à 14 h. T.U. les liaisons par ondes décamétriques entre Paris et New-York ont été totalement impossibles.

On ne peut affirmer ici, qu'il y a une relation certaine de cause à effet entre telle P.I.D.B. et tel trouble ultérieur dans la propagation, mais il est certain que l'on peut dès maintenant envisager l'amélioration de messages d'alerte grâce à une meilleure connaissance des perturbations ionosphériques à début brusque et aux phénomènes qui s'y rattachent. Les enregistrements permanents sont en tout état de cause d'un secours inappréciable.

BIBLIOGRAPHIE

- APPLETON, E. V., 1945, *Nature, Lond.*, 156, 534; 1946, *Phil. Mag.*, 37, 73.
BUREAU, R., 1947, *L'Onde Elect.*, 27, 45.

Diffusion des Echos au Voisinage des Fréquences Critiques de F2

PAR R. RIVAUT

Université de Poitiers

Communicated by J. A. Ratcliffe ; MS. received 26th July 1949. Paper read at Physical Society Conference at Cambridge, 14th-16th July 1949

ABSTRACT. From the study of ionospheric data recorded at Poitiers (46° N.) it appears that two sorts of scattered echoes take place in the neighbourhood of F2 critical frequencies.

The first one seems to be produced by the multiplication of magneto-ionic components. This scattering ends with ground sunrise; it exists during both quiet and perturbed ionospheric conditions; it is far more frequent and stronger when the nights are longer. Its existence seems to be connected with nocturnal thermal and convective disorder.

The second sort of scattering is mainly recorded during the second part of the night, after F2 penetration, at a virtual height of about 800 to 1,200 km. The sharp lower boundary of these reflections, which disappears a few moments after sunrise, leads to the postulation of an ionospheric G region, whose existence has been often discussed. This sort of scattering is connected with magnetically disturbed conditions and, chiefly, with meteoritic periods.

DANS le cas le plus général, on déduit de la théorie magnéto-ionique que, pour un certain niveau d'ionisation maximum de la région F2, trois fréquences critiques peuvent être enregistrées, correspondant aux relations:

$Ne^2/4\pi^2m = f_{N\max}^2 = f_o^2$ pour le rayon ordinaire,

$f_{N\max}^2 = f_{x1}(f_{x1} - f_H)$ pour l'extraordinaire le plus intense,

$f_{N\max}^2 = f_{x2}(f_{x2} + f_H)$ pour l'extraordinaire le moins intense, avec $f_H = eH/2\pi m$.

Donc: $f_{x1} - f_o = f_H f_{x1}/(f_{x1} + f_o) \simeq f_H/2$ si f_o et f_{x1} sont grands devant f_H , et:

Les sondages à la verticale effectués à Poitiers (lat. $46^{\circ} 34' N.$) comprennent habituellement les composantes ordinaire f_o et extraordinaire f_{x1} , espacées d'environ 0,6 Mc/s., soit $f_H/2$. La composante f_{x2} n'apparaît que très rarement, généralement sur les réflexions d'ordre 2 et dans les premières heures de la matinée espacée de f_{x1} d'environ 1,2 Mc/s. (Figure 1*). L'on rencontre beaucoup plus fréquemment des composantes régulièrement ou irrégulièrement réparties dans l'intervalle de fréquence $f_{x1}-f_{x2}$. Chacune des composantes f_o et f_{x1} peut d'abord se dédoubler (Figure 2), puis se multiplier en même temps que les échos perdent leur netteté (Figures 3, 5 et 6), pour, enfin, s'agglomérer en une masse diffuse (Figure 4). Quand ces composantes supplémentaires sont assez fines pour permettre des mesures, on remarque très souvent qu'il existe entre elles des écarts de fréquence privilégiés, de l'ordre de 0,3 (Figure 3) et de 0,15 Mc/s. (Figure 5), valeurs sous multiples de f_H .

A Poitiers, la composante verticale du champ magnétique terrestre est environ le double de la composante horizontale (0,42 et 0,20 gauss au sol), et l'on peut penser que, sur leur trajet, les ondes sont soumises temporairement à des conditions de propagation alternativement transversales ou longitudinales; ceci peut être dû à une variation soit du nombre de chocs ν , provoquée par une modification locale de la densité du milieu ou du nombre d'électrons, soit du champ magnétique terrestre H . Breit a montré, en effet, qu'il suffit d'une variation de H de l'ordre de 20γ pour faire tourner le plan de polarisation de l'onde de 90° . Etant donnée la précision bien inférieure des mesures possibles sur les enregistrements, il n'est pas surprenant que la valeur de f_H , déduite de celles-ci, n'en paraisse pas affectée: les écarts de fréquence des composantes homologues ordinaires et extraordinaires de la figure 5 sont, par exemple, de l'ordre de 0,6 Mc/s. Le rôle de la polarisation est mis en évidence par le fait que ces composantes multiples n'apparaissent pas simultanément sur les échos d'ordre un et ceux d'ordre supérieur; une analyse plus fine nécessiterait la mise en oeuvre de dispositifs expérimentaux de polarisation connue et variable.

L'examen de nombreux enregistrements analogues aux figures précédentes conduit à imaginer la diffusion des échos qui se manifeste à certaines époques, au moment de la pénétration de F2, comme le résultat de la multiplication du nombre des composantes magnéto-ioniques; c'est la même cause physique qui se traduit par ces différents aspects.

De l'étude des cas de diffusion enregistrés de juillet 1948 à avril 1949, entre 3,1 et 11,8 Mc/s., on peut tirer les conclusions suivantes:

(i) la diffusion des échos, au voisinage des fréquences critiques de F2, ne se produit nettement qu'entre le coucher et le lever du soleil;

(ii) la fréquence du phénomène, faible pendant l'été, croît à partir de septembre pour devenir quasi quotidienne en décembre, puis décroît ensuite pour n'être plus qu'accidentelle en mars;

(iii) elle se manifeste généralement dans la deuxième partie de la nuit mais peut débuter dans la première et d'autant plus tôt que les nuits sont plus longues. Elle prend fin au moment du lever du soleil au sol, un peu après le début de l'accroissement matinal des fréquences critiques;

(iv) les enregistrements à caractère diffus sont obtenus aussi bien dans des conditions magnétiques calmes que perturbées (figure 6); le phénomène ne semble pas modifié, non plus, au cours des périodes de grande activité chromosphérique solaire.

* For all photographs see group opposite p. 128.

La diffusion paraît ainsi directement liée à la longueur de la nuit, comme si, entre autres causes possibles, cette durée favorisait le désordre thermique et électronique de l'ionosphère, détruisant la stratification diurne jusqu'au moment du lever solaire.

Il existe une autre sorte de réflexions diffuses, enregistrées à des hauteurs virtuelles de 800 à 1200 kilomètres, sur des fréquences immédiatement supérieures aux fréquences critiques de F2. Elles peuvent se produire au cours de la journée et sont alors faibles (figure 7). Mais elles sont parfois reçues d'une façon intense dans la deuxième partie de la nuit, quand les fréquences critiques de F2 sont basses; leur aspect particulièrement net conduit alors à imaginer l'existence d'une véritable région ionisée désignée par la lettre G (figure 8), existence niée par Bailey, Payne Scott et McCready à la suite d'expériences d'un caractère tout différent.

Ces réflexions nettes se produisent généralement sur des fréquences allant de 3 Mc/s. environ à 5,2 Mc/s.; elles disparaissent, exceptionnellement, sur une fréquence plus élevée (figure 9). Un sondeur plus puissant serait nécessaire pour mettre en évidence les fréquences critiques de cette région. Elles prennent habituellement fin aux environs du lever solaire, comme la diffusion étudiée dans la première partie de cette note. Il arrive, du reste, que ces deux types d'échos diffus se confondent, la région G perdant progressivement son caractère propre pendant que les composantes magnéto-ioniques habituelles s'estompent dans une masse diffuse (figure 10).

Ces deux espèces de diffusion ne sont cependant pas toujours concomitantes, la présence de la région G coïncidant souvent avec des époques magnétiquement perturbées et, surtout, avec le passage des principaux essaims de météorites.

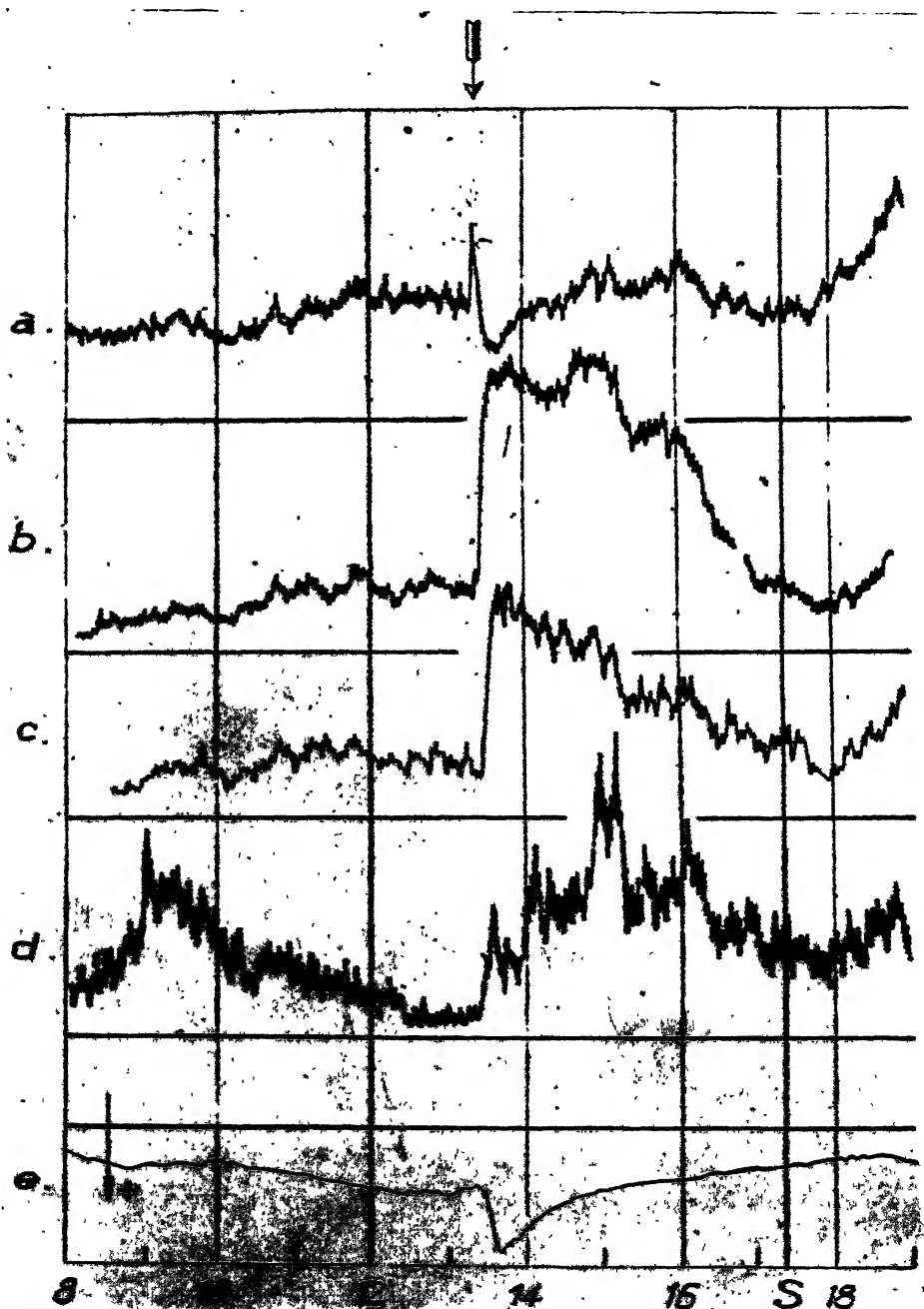


Figure 1. P.I.D.B. du 21 Février 1942.

Successivement à partir du haut :

- a : Enregistrement des atmosphériques sur 23,000 m.
- b, c : Enregistrement des atmosphériques sur 11,000 m. (deux appareils distincts).
- d : Enregistrement des atmosphériques sur 5,000 m.
- e : Enregistrement de la déclinaison magnétique à Chambon la Forêt.

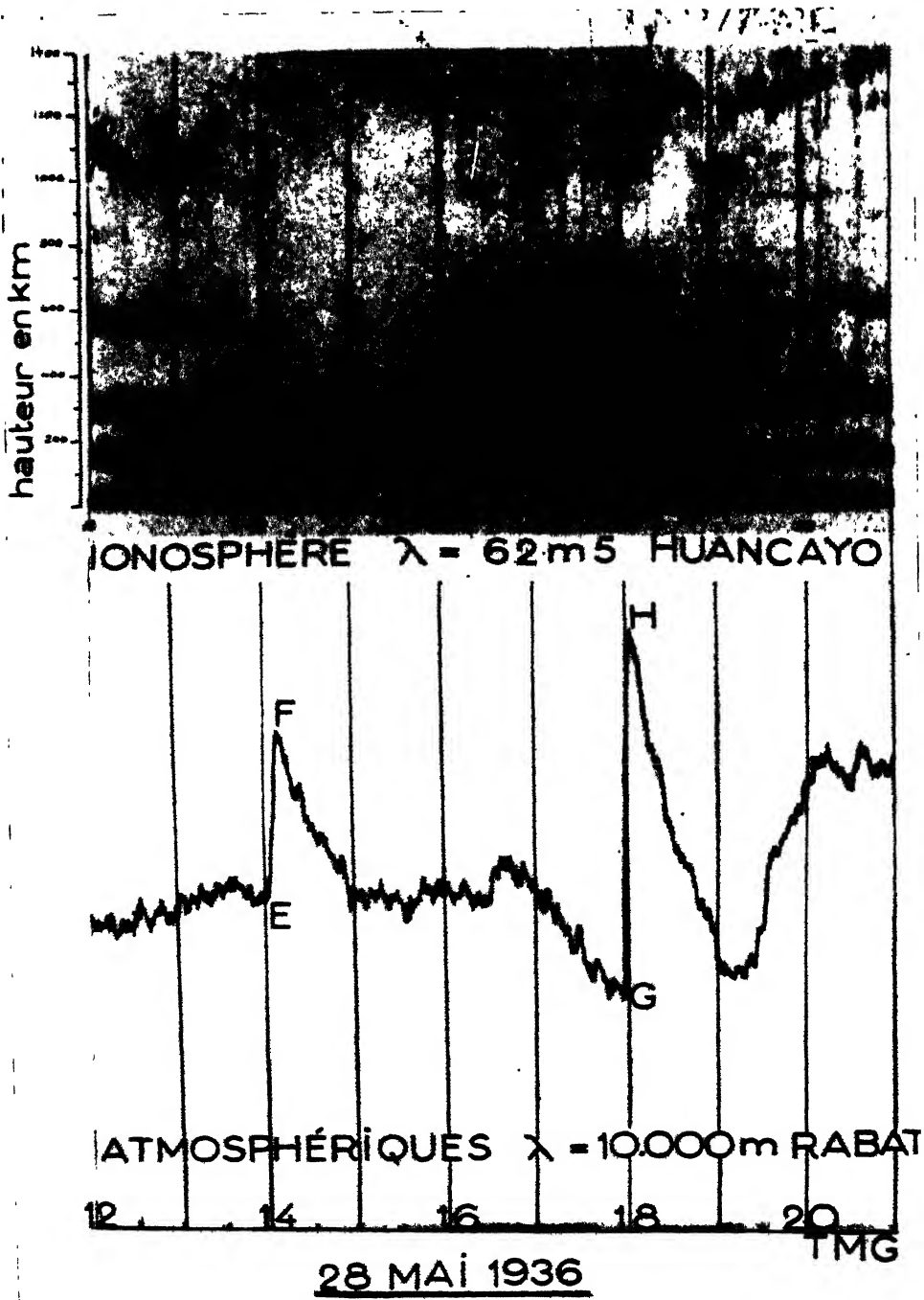


Figure 2.

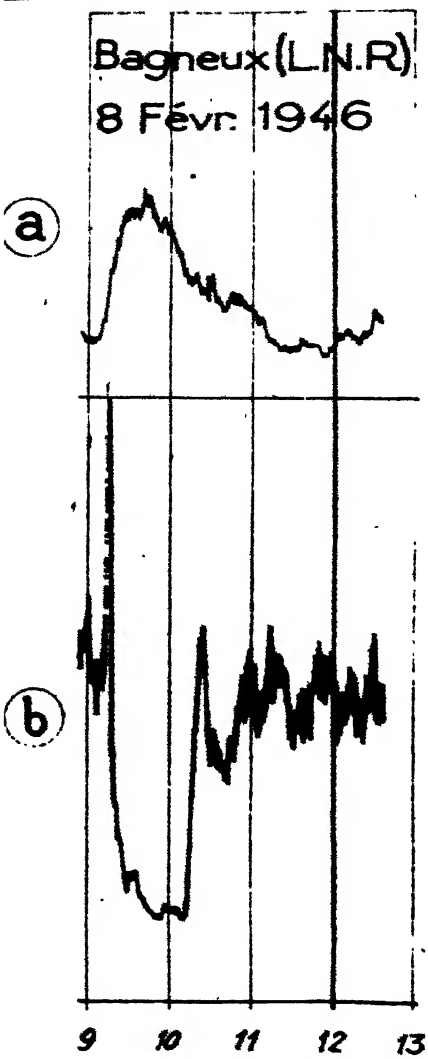


Figure 4.
(a) Atmosphériques 11,000 mètres.
(b) Champ Genève 48,66 mètres.

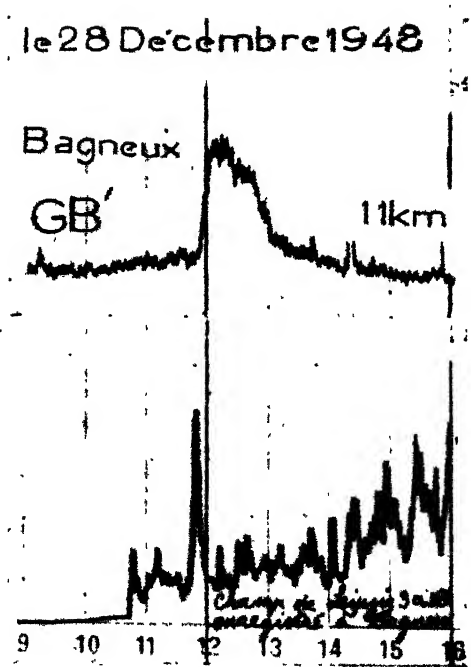


Figure 5.

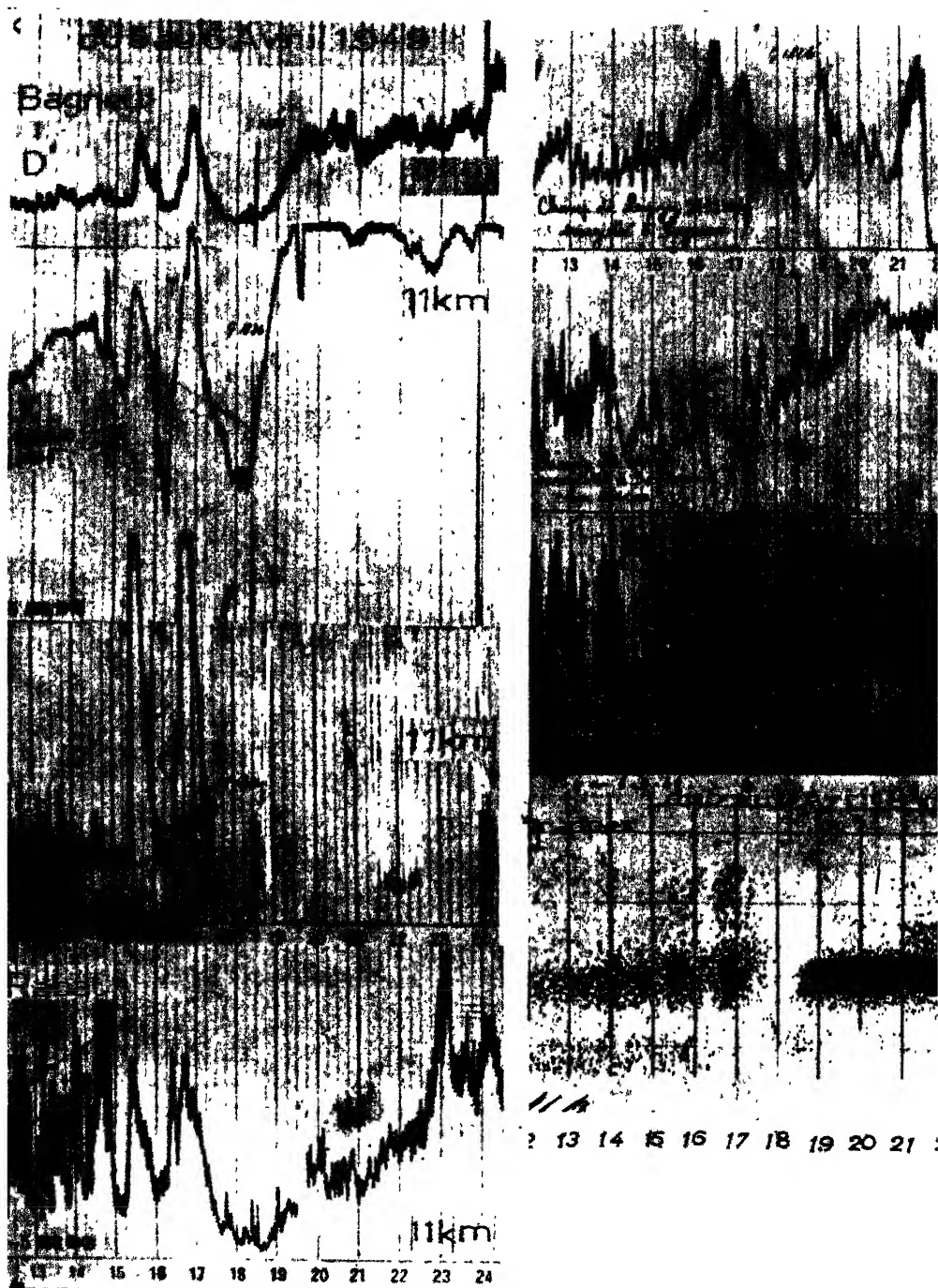


Figure 6.

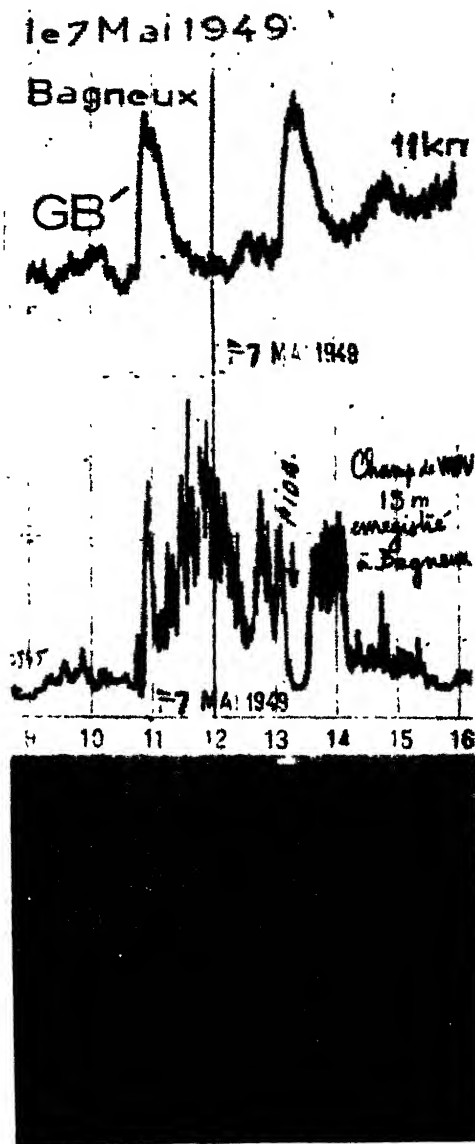


Figure 7.

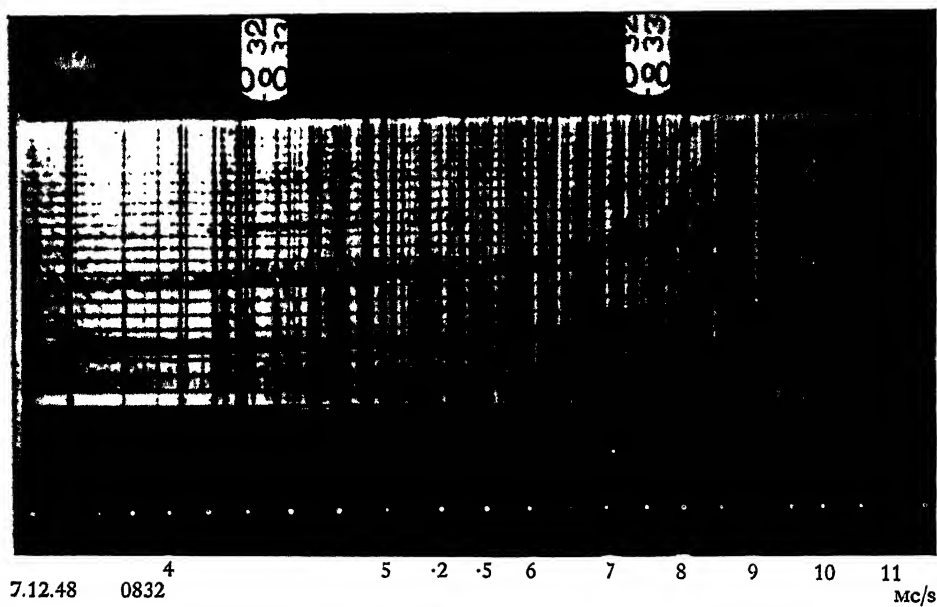


Figure 1.

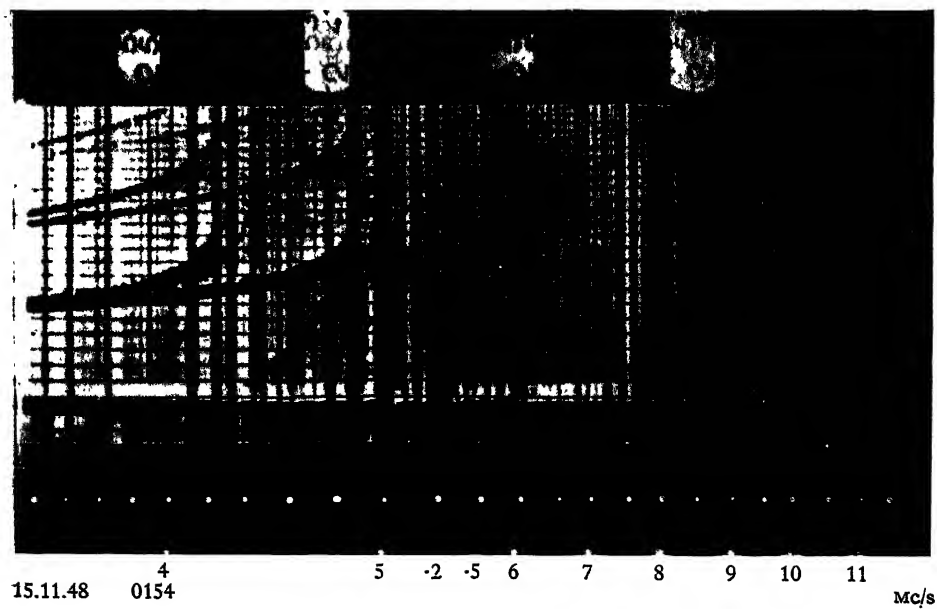


Figure 2.

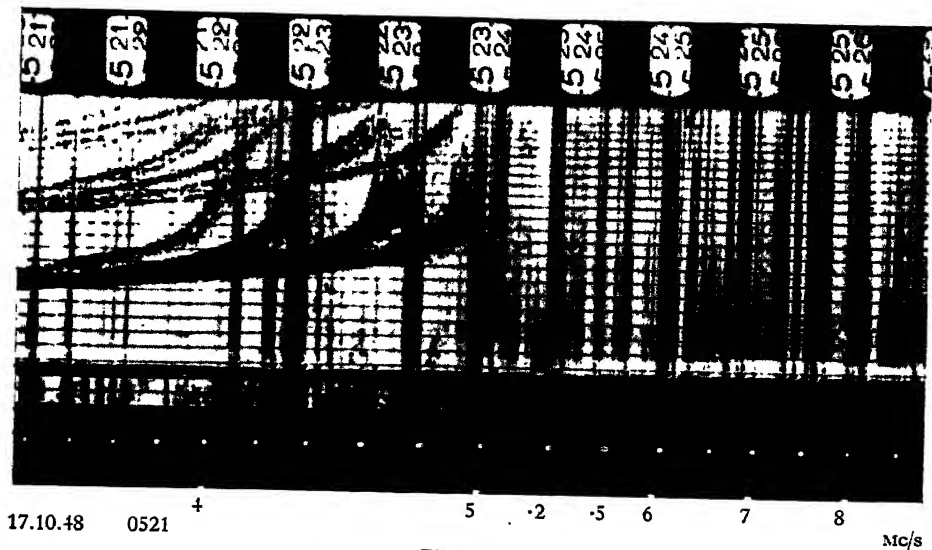


Figure 3.

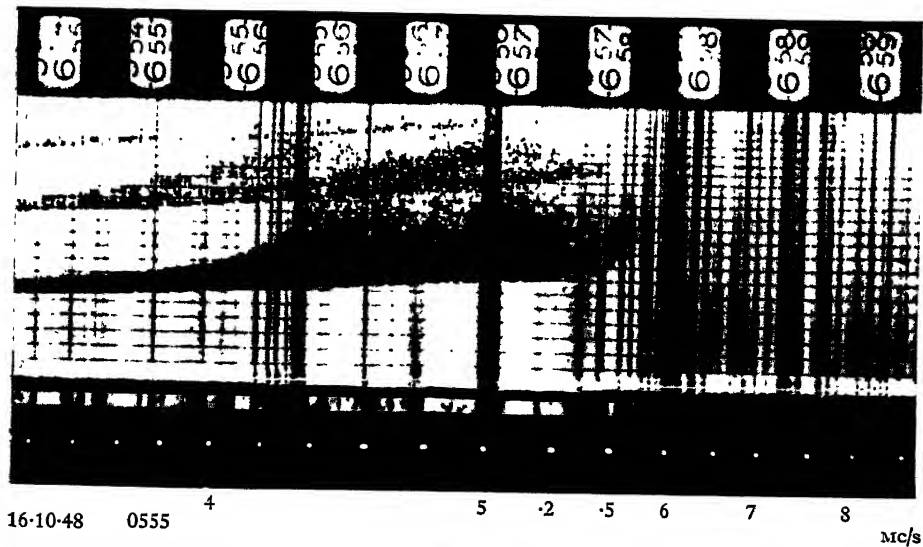


Figure 4.

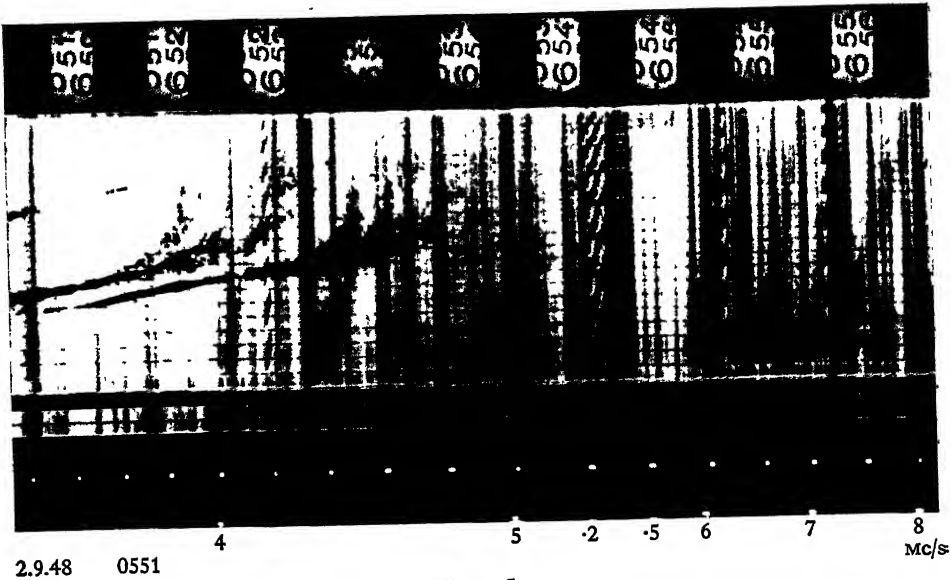


Figure 5.

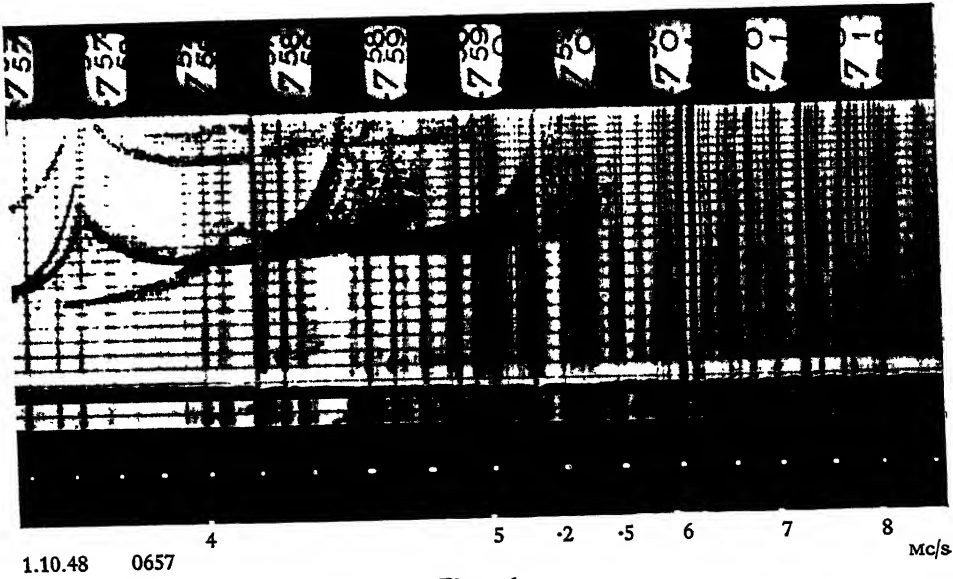


Figure 6.

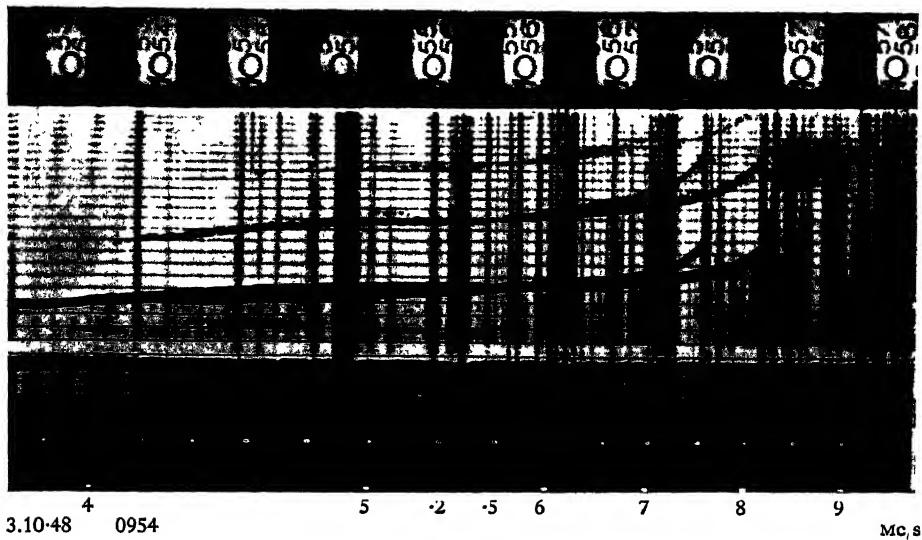


Figure 7.

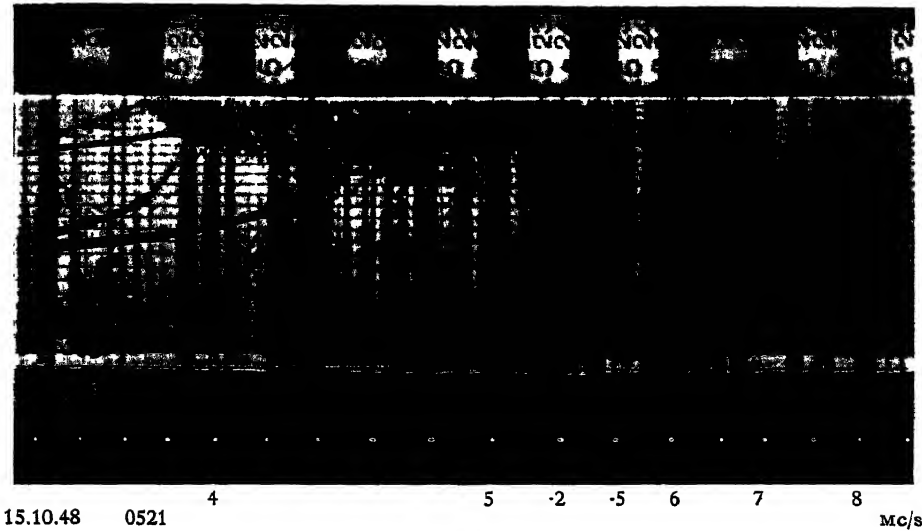


Figure 8.

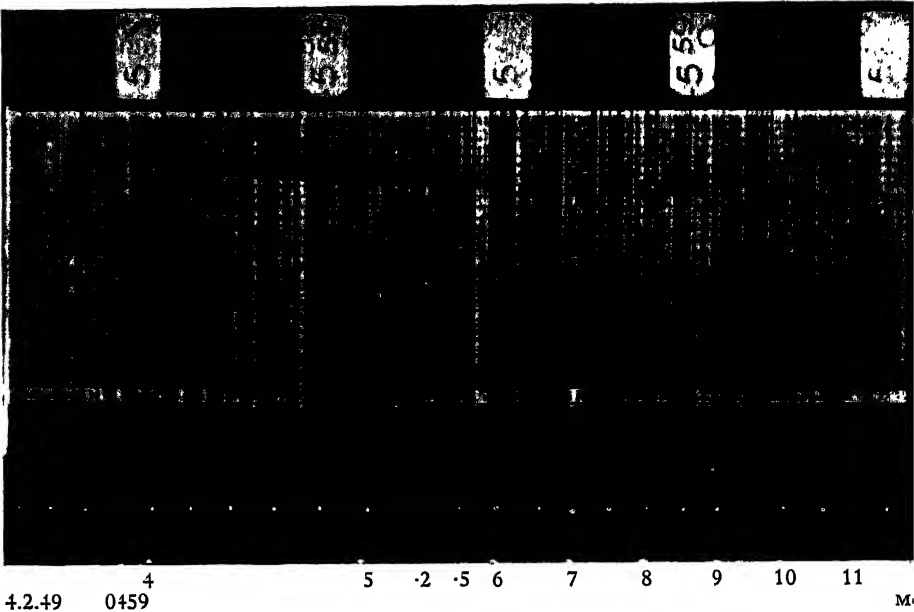
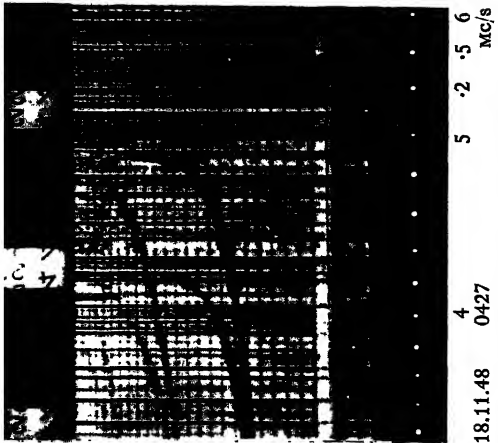
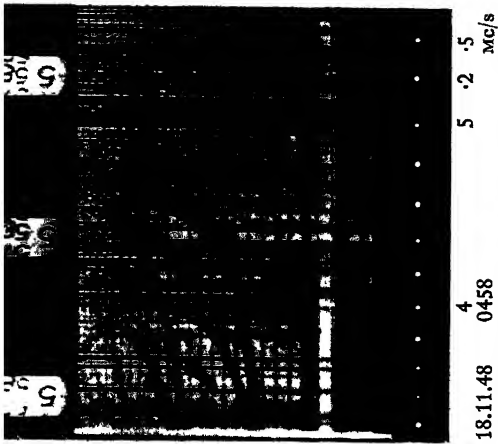
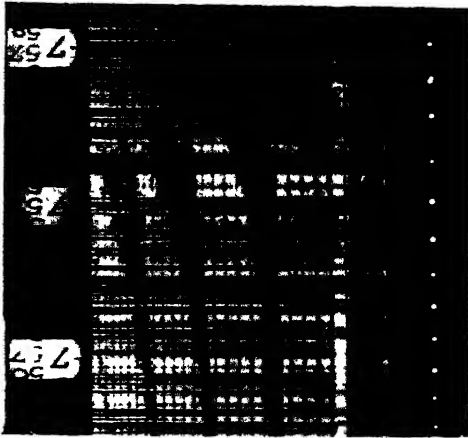
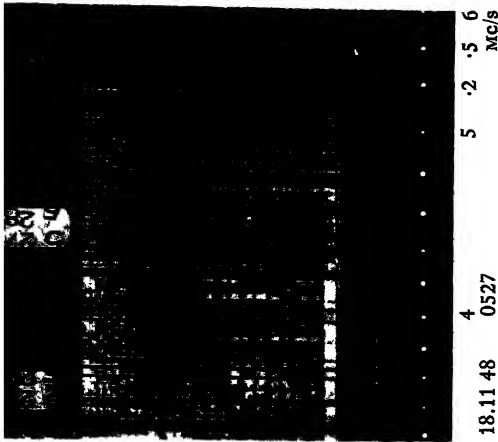


Figure 9.





Theoretical Considerations Regarding the Formation of the Ionized Layers

BY D. R. BATES AND M. J. SEATON

University College, London

MS. received 1st September 1949; read at the Summer Meeting of the Physical Society at Cambridge, July 1949

ABSTRACT. The detailed mechanisms involved in the formation of the E, F₁, F₂ and D layers by solar ultra-violet radiation are discussed. Use is made of the results of some recent calculations on the continuous absorption cross section of atomic oxygen and nitrogen, and of the evidence on the ionic composition of the layers that is provided by the analysis of the emission spectrum of the upper atmosphere during twilight. The uncertainties existing at present are emphasized.

§ 1. INTRODUCTION

FOR a proper theory of the formation of the ionized layers it is necessary to have detailed information on the incident flux of solar radiation, on the altitude distribution and chemical composition of the terrestrial atmosphere, and on the absorption cross-section curve associated with each constituent. Unfortunately the information available on these topics is far from complete and in consequence the development of the theory must proceed slowly and tentatively. In the present paper we confine ourselves to a discussion of the origin of the so-called ultra-violet layers (that is, the E, F₁, F₂ and D layers), and ignore the effects of corpuscles and meteors; for convenience, we treat the three upper, and the lowest, separately. Emphasis is laid on the existing uncertainties.

§ 2. THE E, F₁ AND F₂ LAYERS

2.1. Some years ago Bradbury (1938) suggested that the F₁ and F₂ layers might have a common origin. Bates (1949 a) has recently examined this hypothesis quantitatively and has found that it is consistent with the observational data. Indeed the conclusion that it *must* be largely correct is difficult to avoid as the ionization rate near the 300 km. level due to radiation capable of producing the F₁ layer can be shown to be sufficient *by itself* to account for the F₂ layer also. That there are two layers is simply due to the decrease of the effective recombination coefficient with altitude. The F₁ layer approximates to the standard Chapman type (1931 a) so that its maximum occurs near the level at which the concentration $n(X)$ of the atoms or molecules X responsible for the ionization is given by

$$n(X) = \cos \chi / HA, \quad \dots\dots (1)$$

χ being the solar zenith angle, H the local scale height and A the relevant absorption cross section. In contrast the F₂ layer lies above the region where there is a marked diminution of the intensity of incident radiation and is what may be termed a low attenuation layer, the principal characteristic of which is the relatively weak direct control over the ionization rate exerted by the elevation of the Sun—actually for an extreme zero attenuation layer the ionization rate depends only on the gas density at the level concerned. As far as can be judged the observed diurnal and seasonal variations are in agreement with the theory; analysis of the

irregular fluctuations associated with solar activity shows, too, the expected correlation between the behaviour of the two layers (cf. Allen 1948, Bates 1949 a).

The atmosphere in the region above about 100 km. is generally considered to be composed mainly of molecular (or perhaps atomic) nitrogen and atomic oxygen. It is reasonable therefore to presume that photo-ionization of one or more of these is responsible for the F_1 and F_2 layers. There is some spectroscopic evidence which strongly suggests that the active constituent is not molecular nitrogen. N_2^+ ions resonate strongly in sunlight, emitting bands of the first negative system :



In consequence the emission spectrum of the twilit upper atmosphere provides an extremely sensitive means of detecting their presence. By analysing observational data published by Dufay and Dufay (1947) Bates (1949 a) has shown that during twilight $[N_2^+]$ is only 10 per cm^3 or less at the level of the F layers, which is far smaller than the total ion concentration. The simplest interpretation* is that ionization of molecular nitrogen occurs very slowly and is only of minor importance.

The suggested elimination of molecular nitrogen by the spectroscopic evidence mentioned is particularly useful as otherwise the lack of quantitative information on its absorption cross-section curve would be a serious hindrance to progress. Fortunately both the remaining constituents are atomic, and thus their photo-ionization can be investigated theoretically without great difficulty. Some early calculations on the cross sections were performed by Bates, Buckingham, Massey and Unwin (1939) and by Bates (1939). Adopting the results given by these, and making reasonable estimates of the necessary atmospheric parameters, Bates and Massey (1946), using Chapman's formula (1), found that the location of the F_1 layer appeared to be in fair agreement with what it would be if the layer were produced by the ionization of atomic oxygen,



They also drew attention to the fact that the flux of photons from the conventional 6,000° K. black body model of the Sun is of the order required to account for the observed rate of ionization. It seemed plausible therefore to assume that process (3) is responsible for the F_1 layer; and, as will be recalled, the F_2 layer does not require a different mechanism. The possibility that the ionization of atomic nitrogen,



gives some contribution to both layers is not, of course, excluded.

As far as the E layer is concerned, Bates and Massey (1946) pointed out that if the theory just mentioned were accepted there would be a severe restriction on the processes that could be involved as radiation of quantal energy between 13.5 ev. (the first ionization potential of O) and well over 25 ev. could not penetrate sufficiently deeply into the atmosphere. Two suggestions have been made: Nicolet (1945) has proposed that the responsible mechanism is pre-ionization of molecular oxygen in the energy range 12.2 to 13.5 ev., and Hoyle and Bates (1948) have given consideration to the possibility of general ionization by photons of

* Certain reservations on this will be found in the original paper

very high energy (perhaps about 325 ev.) originating in the solar corona. Both must be regarded as tentative as they depend respectively on the unproven assumptions that pre-ionization of molecular oxygen can occur and that the coronal emission is of sufficient intensity. The observational evidence is still indecisive, but tends to favour the former hypothesis rather than the latter (cf. Bates 1949 a).

It is clear that information on the absorption cross sections of atomic oxygen and nitrogen finds wide application in the detailed discussion of layer formation. In view of this Bates and Seaton (1950) have recently attempted to refine the earlier calculations by including allowance for the effects of electron exchange and by making use of both the dipole moment and dipole velocity formulae for the transition matrix element. The new cross sections obtained are given in the Table. They are appreciably below (actually about half) the old values. While the corrections are not unduly great, their sense, unfortunately, is such as to increase some minor discrepancies already appearing in the provisional theory outlined above. Consequently it is desirable to reconsider the position.

	Ionization potential (ev.)	Absorption cross section at spectral head (A) (10^{-18} cm 2)
Atomic oxygen (O)	13.5	2.6
Atomic nitrogen (N)	14.5	9

The simplest method of studying the significance of the cross sections is to use them in conjunction with formula (1) to determine the atmospheric parameter $Hn(X)$. For the sake of definiteness it is convenient to adopt a specific value of $\cos \chi$; 0.8 is a sufficiently representative figure.

Consider the ionization of atomic oxygen by photons in the energy range 13.5 to 16.9 ev. As can be seen at once* from the Table, $Hn(O)$ is about $3 \times 10^{17}/\text{cm}^2$. Now at the level of the F_1 layer H and $n(O)$ are usually assumed to be 30 km. and $3 \times 10^{10}/\text{cm}^3$, respectively †, giving a value of $Hn(O)$ equal to $1 \times 10^{17}/\text{cm}^2$, which is three times smaller than the theory demands. It is difficult to decide whether or not the discrepancy is real. The accuracy of the atmospheric data is very uncertain ‡, and it may well be that the conflict is caused by the combination of errors in H and $n(O)$, and perhaps to some extent from a decrease of the recombination coefficient with altitude (which would tend to raise the maximum of the layer). But to obtain agreement it is necessary to postulate *greater* values of the first two quantities than are normally assumed. A change of this nature in the supposed structure of the atmosphere, while not inconceivable, cannot be accepted lightly, involving as it does an increase in the already high temperature. It is worth considering therefore whether there is any possibility that atomic oxygen is *not* the constituent responsible for the F_1 layer.

* Molecular nitrogen is so transparent in the spectral region concerned that no correction need be made for absorption by it. We are grateful to Dr. W. C. Price and Dr. R. E. Worley for information in this connection.

† The adoption of the A appropriate to the spectral head is only justified if the flux of the solar radiation decreases rapidly with energy (as is the case for a $6,000^\circ \text{K}$. black body). If the fall off proves to be slow, a higher mean A should be taken to allow for ionization to $O^+(^2D)$ and $O^+(^2P)$. The effect of this would be to decrease the discrepancy mentioned in the text.

‡ The data used here (and elsewhere in the paper) are taken from the 'model atmosphere' proposed by Bates and Massey (1946). The basis for the estimates is further discussed in a recent review (Bates 1949 b).

§ Harang (1945), for example, advocates scale heights and densities that are rather different from those that have been adopted.

On account of the intensity of the ionization likely to result from process (3) it is reasonable to suppose that its maximum is located in some layer. As an alternative to the F_1 layer the E layer might be suggested. In it H is about 10 km. and $n(O)$ is about $2 \times 10^{12}/\text{cm}^3$, so that $Hn(O)$ is approximately $2 \times 10^{18}/\text{cm}^2$. The theoretical figure given in the previous paragraph must not be compared directly with this. It is necessary to make allowance for the probable rapid upward increase of the scale height in the region concerned. Assuming for simplicity that it is a linear function of the altitude h , Chapman's formula is replaced by

$$n(X) = (1 + dH/dh) \cos \chi / HA, \quad \dots\dots(5)$$

the H in the denominator being that at the layer maximum. dH/dh may be as great as unity, so that the value of $Hn(O)$ based on the calculated cross section now becomes perhaps $6 \times 10^{17}/\text{cm}^2$. Though the discrepancy is thus no larger than for the F_1 layer it is rather less easy to ignore as H and $n(O)$ are better determined in the E layer; further, H can scarcely be below the 10 km. adopted so that the entire error would have to be attributed to $n(O)$, which would hence need to be $6 \times 10^{11}/\text{cm}^3$ instead of the estimated $2 \times 10^{12}/\text{cm}^3$. The E_2 layer, being slightly above the normal E (or E_1) layer, is clearly more satisfactory as far as the location requirement is concerned. However, in view of the fact that it can only be observed occasionally, the possibility that an ionizing action as effective as that under discussion is involved must be treated with caution. Radio evidence on the factors governing its appearance, and on the density of ionization in the region between the two main layers, would be useful.

To complete our exploratory enquiry we must examine whether any process other than (3) is capable of producing the F_1 and F_2 layers: for unless there is some such process the various suggestions put forward for consideration must be abandoned. The ionization of atomic nitrogen at once presents itself as a possible mechanism. Taking H to be 30 km. (as before) the value of $n(N)$ that is required in the Chapman-like F_1 layer can be seen to be $3 \times 10^{10}/\text{cm}^3$ (compared with a total particle concentration of $1 \times 10^{11}/\text{cm}^3$). Though this figure would imply a high degree of dissociation it cannot be said to be impossible (cf. Bates 1949 b). The integrated rate of ionization/ cm^2 column in the F layers is, of course, several times greater than that in the E layers. It might be thought at first that this is in serious conflict with the fact that the ionization potential of N is higher than that of O (14.5 ev. compared with 13.5 ev.). If the $6,000^\circ\text{K}$. black body model of the Sun is used the difficulty certainly appears very acute. However, such a model, though providing a useful general guide for some purposes, does not necessarily correspond with reality; and indeed recent work by Hoyle* (1949) suggests that the intensity of the solar radiation beyond the Lyman limit falls off so slowly with increasing frequency that the photon flux above 14.5 ev. is actually greater than that between 13.5 and 14.5 ev.

2.2. In summarizing the position it is convenient to regard the development of the theory of the formation of the ionized layers as progressing in two main stages: (a) the listing of all processes that might conceivably have to be taken into consideration, and the investigation of which combinations of them might give a plausible account of the observed ionospheric structure; and (b) the

* We wish to thank Mr. Hoyle for showing us his results before publication and for many stimulating discussions.

detailed quantitative examination of the alternatives to decide which is correct. It seems probable that the speculative but essential first stage is largely completed. However, though the conditions that must be satisfied are fully appreciated, and though some important restrictions on the possibilities are now realized, much of the second and far more difficult stage has still to be traversed. In order to discriminate unambiguously between the theories suggested it is necessary to have precise information on certain atmospheric and other parameters. The task of attaining the required degree of accuracy is formidable. Until it is accomplished the exclusive advocacy of a particular theory is unjustified. In particular no firm decision can yet be reached as to whether the F layers arise mainly from atomic oxygen, or whether the maximum of the ionization rate due to this constituent occurs at a lower level and the principal active constituent responsible for the F layers is atomic nitrogen.

§3. THE D LAYER

3.1. The theories of the origin of the D layer cannot be properly discussed without some knowledge of the rate of ionization to be explained. Elaborate calculations are not merited at present and we will therefore make estimates of this rate only at two specific levels, the 75 km. and 90 km. levels. As these are believed to be near the base and near the maximum of the layer, respectively, the salient features will be illustrated adequately. It must be stressed that the figures we deduce are essentially of a preliminary nature and will doubtless have to be revised later. Our intention in giving them is merely to provide an indication of the problem that is presented.

Radio scientists have published very little quantitative data on D layer ionization. However Mr. W. R. Piggott has been generous enough to acquaint us with the results he has obtained during the past few years from absorption measurements. It appears probable from his investigations that at the 90 km. level the electron concentration, $n(e)$, is about $1.5 \times 10^4/\text{cm}^3$ at mid-day, and the effective recombination coefficient α is about $3 \times 10^{-8} \text{cm}^3/\text{sec}$. In the near equilibrium conditions prevailing the effective rate of electron production, q , may be taken as equal to $\alpha n(e)^2$. From the figures just quoted its value can be seen to be approximately $7/\text{cm}^3/\text{sec}$.

The lower part of the layer is best explored by the study of the reflection of long radio waves. Dr. K. Weekes has informed us that the most recent work using this technique indicates that $n(e)$ is some $2.5 \times 10^3/\text{cm}^3$ at or near the 75 km. level at mid-day. While α in this region has not as yet been determined, some provisional information on it can be obtained: thus by using the theoretical ionization curves compiled by Chapman (1931 b) it can readily be shown that in order that there be sufficient time for $n(e)$ to reach the observed value, q at noon must be at least some $6 \times 10^{-2}/\text{cm}^3/\text{sec}$., which implies that α must be at least of order $1 \times 10^{-8}/\text{cm}^3/\text{sec}$.

Now α and q are not themselves of fundamental significance: they are simply parameters appearing in the familiar equation governing the variation of $n(e)$ with time:

$$dn(e)/dt = q - \alpha n(e)^2. \quad \dots\dots(6)$$

For our purpose we are interested in the actual electronic and ionic recombination coefficients, α_e and α_i respectively, and in the true rate of photo-ionization, $\mu n(X)$.

The two sets of quantities are related. Introducing λ to represent the ratio of the concentration of negative ions to that of electrons it may be proved that

$$\alpha \simeq \alpha_e + \lambda \alpha_i \quad \dots\dots (7)$$

and

$$\mu n(X) = (1 + \lambda)q \quad \dots\dots (8)$$

(Appleton and Sayers 1938, Bates and Massey 1946). Formula (8) is of particular importance in showing that q is less than $\mu n(X)$. The figures given in the previous paragraphs must therefore only be regarded as setting a lower limit to the photo-ionization rate.

An estimate of λ can be made by studying the equilibrium between the reactions leading to the formation and destruction of negative ions. On account of the relatively high gas densities involved it is probable that electrons become attached to oxygen molecules mainly through the Block-Bradbury process (Bates and Massey 1946). The coefficient η associated with this depends on the concentration of molecular oxygen: analysis of the scanty experimental data available indicates that its value may be some 2.5×10^{-14} cm³/sec. at 75 km. (where $n(\text{O}_2)$ is taken as 4×10^{14} /cm³) and some 1×10^{-14} cm³/sec. at 90 km. (where $n(\text{O}_2)$ is taken as 8×10^{13} /cm³). During the day the O_2^- ions formed must suffer photo-detachment,



Crude calculations suggest that the rate ρ at which this occurs may be perhaps about 0.5 per negative ion per second. Hence from the formula

$$\lambda \simeq \eta n(\text{O}_2)/\rho \quad \dots\dots (10)$$

it can be seen that

$$\lambda_{75 \text{ km.}} \simeq 20 \quad \dots\dots (11)$$

$$\lambda_{90 \text{ km.}} \simeq 2. * \quad \dots\dots (12)$$

Unfortunately the uncertainties in η and ρ are so serious that these figures should only be regarded as indicating the possible order of magnitude at each of the two levels concerned. Equation (7) however provides at least some confirmation that $\lambda_{75 \text{ km.}}$ is large, for otherwise it is difficult to explain an α of 10^{-8} cm³/sec. since neither α_e nor α_i are likely to be much in excess of 10^{-8} cm³/sec. (Bates and Massey 1946, 1947).

Finally substitution in the expression for the photo-ionization rate (8) yields

$$\mu n(X)_{75 \text{ km.}} \simeq 1.3/\text{cm}^3/\text{sec.}, \quad \dots\dots (13)$$

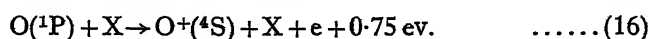
$$\mu n(X)_{90 \text{ km.}} \simeq 20/\text{cm}^3/\text{sec.} \quad \dots\dots (14)$$

The fact that these estimates may be in error by a considerable factor must be kept constantly in mind throughout the remainder of the discussion.

3.2. A number of theories of D layer ionization (normal and enhanced) have been put forward but have not been developed sufficiently quantitatively for their merits and defects to be properly assessed. For reference purposes we will first summarize them briefly.

* It may be noted that below the 75 km. level λ increases rapidly downwards, probably indeed as the square of the gas density. The effective recombination coefficient α must increase at least at the same rate (or faster if three body processes contribute to α_i) and must attain a very large value in the region immediately beneath the base of the normal layer.

In 1936 Chapman and Price drew attention to the close coincidence that exists between $\lambda 1215.6 \text{ \AA}$, the wavelength of the $L(\alpha)$ line of hydrogen, and $\lambda 1217.6 \text{ \AA}$, the wavelength of the radiation required to raise atomic oxygen from the metastable state, $(p^4)^1S$, to the highly excited state, $(p^3s)^1P$; they also pointed out that an atom in this latter state might suffer ionization as a result of a collision of the second kind. Remarking that the solar spectrum may contain $L(\alpha)$ as a broad emission line they suggested that appreciable atmospheric ionization might be produced by the sequence



Later Martyn, Munroe, Higgs and Williams (1937) proposed that this mechanism might be responsible for the D layer.* Other authors have favoured ordinary photo-ionization processes (Mitra, Bhar and Ghosh 1939, Nicolet 1945, Bates and Massey 1946). The more obvious possibilities are:



These possibilities have one important characteristic in common—they involve radiation of comparatively high photon energy.† Jouaust and Vassy (1941) and later Vassy and Vassy (1942) have claimed that laboratory experiments indicate that such radiation (and indeed radiation of photon energy down to about 7 ev.) is so strongly absorbed by air that it is impossible for it to penetrate to the required levels. To avoid this apparent difficulty they suggested that the mechanism operative might be the photo-ionization of atomic sodium,



the threshold energy for which is only 5.1 ev.

To differentiate between the various theories outlined it is necessary to examine each critically. We will begin by considering the sodium theory, as in view of the remarks in the previous paragraph it might seem to be the only one tenable. $\mu n(Na)$, the contribution to the total ionization rate from (20), can be calculated at least approximately. It depends on $n(Na)$, the concentration of atomic sodium, A , the absorption cross section associated with the process, and S , the incident photon flux in the relevant energy range. Information on all three is available.

(i) Observations on the twilight enhancement of the D lines enable $n(Na)$ to be determined. The most recent work (that of Barbier (1948)) gives a value of approximately $3 \times 10^8/\text{cm}^3$ at 75 km. and $6 \times 10^8/\text{cm}^3$ at 90 km. Variations of course occur—concentrations much smaller than those quoted have been reported.

* The original theory referred to the fade-out enhancement, but for completeness we will consider it extended to include the normal ionization.

† $L(\alpha)$ corresponds to 10.1 ev. The ionization potentials of O_2 and NO are known to be 12.2 and 9.1 ev.; that of O_3 has not been measured, but the theoretical work of Mulliken (1942) suggests it to be 12.5 ev. (which is consistent with the estimated value obtained from the standard dissociation-ionization cycle). It may be noted that the ionization potentials of the other likely constituents are also high (H_2O , 12.6 ev.; N_2O , 12.7 ev.; CO_2 , 13.7 ev.; CO , 13.9 ev.; H_2 , 15.4 ev.).

(ii) Recent measurements on A (Ditchburn and Jutsum, unpublished *) show that at the ionization limit A equals $1.16 \times 10^{-19} \text{ cm}^2$; immediately above it falls off rapidly with increasing frequency.

(iii) By means of rocket-borne instruments the solar spectrum in the main region of importance ($\lambda 2415 \text{ \AA}$. to $\lambda 2200 \text{ \AA}$.) has now been surveyed. Preliminary curves showing the intensity distribution are given in the reports of the U.S. Naval Research Laboratory (1948). The mean absolute value is about one-tenth that from a $6,000^\circ \text{K}$. black body.

Combining the data from these sources it can readily be shown that

$$\mu n(\text{Na})_{75 \text{ km.}} \simeq 3 \times 10^{-2} / \text{cm}^3 / \text{sec.} \dagger \quad \dots\dots (21)$$

$$\text{and} \quad \mu n(\text{Na})_{90 \text{ km.}} \simeq 5 \times 10^{-3} / \text{cm}^3 / \text{sec.} \dagger \quad \dots\dots (22)$$

These rates are very small compared with the rates apparently required (§ 3.1). Though (20) is probably augmented by similar processes involving other metallic elements (traces of which are presumably present in the upper atmosphere) it would seem most unlikely that the effect could be sufficient to make the total contribution appreciable except perhaps at the extreme base of the layer. A further discrepancy also arises. Since the active radiation is not appreciably attenuated, the ionization rate during the day would be expected to be essentially independent of the solar zenith angle. This expectation is in direct conflict with the observations of radio workers. While these are consistent with the existence of a small component which remains constant during the day (cf. Gockel 1948) they show unmistakably that the main ionization rate is closely controlled by the elevation of the Sun (Budden, Ratcliffe and Wilkes 1939). It would seem then that some additional mechanism must be sought.

Before proceeding further we must reconsider the claim of Jouaust, Vassy and Vassy that only low energy photons can penetrate to the level of the D layer. This was based mainly on measurements made by Schneider (1940) on the absorption by air in the extreme ultra-violet ($\lambda 380 \text{ \AA}$. to $\lambda 1600 \text{ \AA}$.). These certainly indicate that photons of energy greater than about 12 ev. are very effectively absorbed. Hopfield (1946) has recently published evidence confirming this. His work is particularly valuable as (unlike Schneider) he used a *continuous* light source. The spectrograms he gives show that the radiation, in the limited range under discussion, is almost completely extinguished in passing through a 0.4 cm. column of air at S.T.P. Now the equivalent path length down to the 90 km. level is about 10 cm. Hence, as can readily be verified, the residual intensity even here is at most only some 10^{-11} times the incident intensity. It is apparent therefore that processes (17) and (18) cannot yield appreciable ionization: the necessary detailed calculations are elementary and need scarcely be presented.

The situation with regard to photons of energy below 12 ev. is less simple. Jouaust, Vassy and Vassy adopted, without reservation, the absorption cross sections tabulated by Schneider. However the figures given are not uniformly reliable. Schneider himself states that those relating to the more transparent

* We wish to thank Professor R. W. Ditchburn for placing at our disposal these results obtained in his laboratory.

† Rates previously estimated (Bates 1947) were greater owing to the adoption of the absorption cross section computed by Rudkjöbing (1940), which is rather larger than that of Ditchburn and Jutsum, and to the acceptance of the view then widely held that the solar emission in the spectral region concerned was somewhat in excess of that from a black body at $6,000^\circ \text{K}$. The values originally given for the ratio $[\text{Na}^+]/[\text{Na}]$ are thus too great, but even with the new data the degree of ionization of atmospheric sodium remains remarkably high.

spectral regions should be regarded only as approximate *upper limits*. One of these regions is of special interest as it includes $L(\alpha)$. Williams (1940) and Preston (1940) have independently studied the absorption of this individual line. Their results did not agree—a large cross section was obtained by Williams and a small cross section by Preston. Jouaust, Vassy and Vassy argued that the former result should be accepted (with minor modifications) and that the latter should be rejected; the accord this gave with their interpretation of Schneider's work seemed reasonable. We have however been in communication with Dr. Williams who has informed us that he has performed further measurements and now considers that Preston's result is correct—that is, that the absorption cross section of $L(\alpha)$ in air is only about $2 \times 10^{-21} \text{ cm}^2$. The extremely low attenuation rate this implies is of great significance. It means that the line penetrates deep into the atmosphere: to reduce the intensity by a factor of 10 an equivalent path of some 42 cm. must be traversed. Doubtless rocket experiments will soon provide conclusive evidence on the matter.

Hopfield's investigation gives no information regarding $L(\alpha)$ as atomic hydrogen was present in the apparatus used. It does however indicate that air has several other 'windows' of high transparency in the photon energy range between 10 and 11 ev. As far as can be judged from the spectrograms published the radiation in these is not seriously attenuated in a path 4.2 cm. long. This suggests that the absorption cross section is not more than some 10^{-20} cm^2 (perhaps about $5 \times 10^{-21} \text{ cm}^2$).

We are now in a position to discuss the two remaining theories of the origin of the D layer.

In estimating the ionization rate $\mu n(\text{O})$ resulting from (15) and (16) it is convenient to express the $L(\alpha)$ intensity emanating from the Sun as x times the intensity of the continuum (at the corresponding frequency ν) emitted by a black body at a temperature T . Before reaching the Earth the radiation is diluted by a factor f , and before penetrating to the D layer it is attenuated by a further factor g . If ω_A and ω_B are the statistical weights of the initial state A, and final state B, of the oxygen atom, if τ is the ratio of the number of spontaneous transitions from B to A to the total number from B to any state, and if ζ is the rate coefficient associated with the ionizing collisions postulated, then it can be seen that

$$\mu n(\text{O}) = \left\{ \frac{\omega_B}{\omega_A} x f g \exp(-h\nu/kT) \right\} \tau \zeta n(\text{X}) n(\text{O}^1\text{S}), \quad \dots\dots(23)$$

k being Boltzmann's constant. The values of all the quantities except x appearing in this formula are either known or can be limited in some way.

We will first summarize briefly the information available, and then consider what are the demands on x made by the theory. The values of ω_A and ω_B are 1 and 3 respectively. Because of the presence of the adjustable factor x we may assign to T the conventional value $6,000^\circ \text{K}$. f is 5.4×10^{-6} (a standard figure). From Preston's measurements on absorption it can be shown that $g_{75 \text{ km.}}$ is of order 6×10^{-2} and that $g_{90 \text{ km.}}$ is approximately 5×10^{-1} . Estimates of the relative transition probabilities suggest that τ is about 0.3 (and in any event it must be less than unity). Since the ionizing collisions violate the spin conservation rule their efficiency would not be expected to be high and it therefore seems improbable that ζ is greater than $1 \times 10^{-11} \text{ cm}^3/\text{sec}$. X, the constituent involved in the ionizing collisions, is unidentified, but a sufficient upper limit to $n(\text{X})$ is provided by N ,

the total particle concentration, which at 75 km. is $2 \times 10^{15}/\text{cm}^3$ and at 90 km. is $4 \times 10^{14}/\text{cm}^3$. A tentative estimation of $n(\text{O}^1\text{S})$ can be obtained by considering the mechanisms leading to the formation and destruction of metastable oxygen atoms. Bates and Massey (1946) have studied the equilibrium between the processes



and have shown that it gives

$$n(\text{O}^1\text{S}) \simeq 3 \times 10^{-11} n(\text{O}^3\text{P}). \quad \dots\dots (25)$$

Detailed calculations have recently been made by Penndorf (1949) on the amount of atomic oxygen in the upper atmosphere. These indicate that $n(\text{O}^3\text{P}_{90 \text{ km.}})$ is unlikely to exceed about $10^{13}/\text{cm}^3$ (and it may well be much smaller). $n(\text{O}^3\text{P}_{75 \text{ km.}})$ was not computed but it must clearly be less than the figure just quoted; $4 \times 10^{12}/\text{cm}^3$ seems a reasonable upper limit to take. Hence from (25) $n(\text{O}^1\text{S}_{75 \text{ km.}})$ is at most $120/\text{cm}^3$ and $n(\text{O}^1\text{S}_{90 \text{ km.}})$ is at most $300/\text{cm}^3$.* Finally, numerical substitution in (23) yields

$$\mu n(\text{O})_{75 \text{ km.}} < 2 \times 10^{-9} x / \text{cm}^3 / \text{sec.} \quad \dots\dots (26)$$

and

$$\mu n(\text{O})_{90 \text{ km.}} < 1 \times 10^{-8} x / \text{cm}^3 / \text{sec.} \quad \dots\dots (27)$$

In spite of the uncertainty with regard to x , these ionization rates appear so slow that the possibility that processes (15) and (16) might account for the ordinary D layer would seem to be excluded. Many authors consider that the intensity of $L(\alpha)$ becomes extremely high during the chromospheric eruptions with which radio fade-outs are associated and that x may actually be of order 10^6 for a major disturbance (Hunter 1942-43). A still greater intensity is required to explain the enhancement of the D layer, so that even for this special phenomenon the theory shows little promise.

It is not unlikely that nitric oxide is an important constituent of the upper atmosphere. Price (1942-43), for example, has pointed out that it is one of the products of dissociation of nitrous oxide which is probably present in considerable concentrations (Sutherland and Callendar 1942-3). There is some reason too for believing that nitrogen as well as oxygen occurs partly in the free atomic form (Bates 1949 b); if this is the case it would be expected that nitric oxide would be formed by chemical action. Again Durand, Oberly and Tousey (1949) have recently reported what may prove to be direct confirmatory evidence. In the course of the analysis of a spectrum of the Sun taken during a rocket flight to an altitude of 55 km. they discovered a broad band lying between $\lambda 2200 \text{ \AA.}$ and $\lambda 2300 \text{ \AA.}$; they suggested that it may be caused by absorption by atmospheric nitric oxide. If this interpretation is correct the number of NO molecules above the 55 km. level must be of order $10^{18}/\text{cm}^2$ column. Nothing can be said at present about the distribution, but it is at least possible that in the region concerned it is much the same as that of the main gases.

Lack of basic data makes it impossible to investigate the photo-ionization of nitric oxide sufficiently quantitatively to establish whether or not it is responsible for the formation of the D layer, but in view of the unsatisfactory position

* We should perhaps mention that Chapman's association-excitation process $3\text{O} \rightarrow \text{O}_3 + \text{O}^1\text{S}$ does not lead to greater concentrations than those given. This can readily be demonstrated by a simple argument based on the known rate of emission of $\lambda 577 \text{ \AA.}$ during the night (when the amount of atomic oxygen is not sensibly less than during the day (Penndorf 1949)). The neglected deactivation collisions will, of course, tend to reduce $n(\text{O}^1\text{S})$.

of the alternative theories considered the possibility appears to be at least worth examining. A calculation so speculative does not justify detailed presentation, and we state briefly the assumptions made and the conclusions reached: (i) Consistent with a total NO content of $10^{18}/\text{cm}^2$ column we take $n(\text{NO})_{75 \text{ km.}}$ to be $2 \times 10^{11}/\text{cm}^3$ and $n(\text{NO})_{90 \text{ km.}}$ to be $4 \times 10^{10}/\text{cm}^3$. (ii) It is known that the cross section associated with process (19) is low (Price 1942-43): we arbitrarily adopt a value of $5 \times 10^{-20} \text{ cm}^2$. (iii) In the absence of more exact information we regard the quiescent Sun as a $6,000^\circ \text{K.}$ black body. Accepting Preston's (1940) and Hopfield's (1946) work we treat the atmosphere as opaque except for two groups of windows: in one (containing $L(\alpha)$ and of width perhaps 5 Å.) the absorption cross section is about $2 \times 10^{-21} \text{ cm}^2$, in the other (of width perhaps 20 Å.) the absorption cross section is $5 \times 10^{-21} \text{ cm}^2$. This gives the photon flux at 75 km. as about $1 \times 10^9/\text{cm}^2/\text{sec.}$ and that at 90 km. about $2 \times 10^{10}/\text{cm}^2/\text{sec.}$

Combining (i), (ii) and (iii) it can be seen at once that

$$\mu n(\text{NO})_{75 \text{ km.}} \simeq 10/\text{cm}^3/\text{sec.}$$

$$\mu n(\text{NO})_{90 \text{ km.}} \simeq 40/\text{cm}^3/\text{sec.}$$

These rates are considerable, and it may be that the photo-ionization of nitric oxide actually does yield the major part of the ordinary D layer. As need scarcely be emphasized, the estimates made are too crude to justify an assured statement.

The nitric oxide theory appears quite attractive as far as the enhancement during fade-outs is concerned, for if $L(\alpha)$ is greatly intensified a high ionization rate would certainly occur unless the values given for $n(\text{NO})$ are grossly in excess. It is interesting to note that on account of the difference in the transparency of the two groups of 'windows' in the atmosphere an intensification of $L(\alpha)$ would tend to cause a lowering of the layer. Radio measurements give some evidence of such an effect (Budden, Ratcliffe and Wilkes 1939).

Mention must, however, be made of a possible difficulty associated with any purely photo-ionization theory of fade-outs. Suppose that the height of reflection of long radio waves is normally h_N , and that it is reduced to h_F during a fade-out. τ , the time for recovery to $\frac{1}{2}(h_N + h_F)$, is simply the time taken for the electron concentration at this mean level to fall from its enhanced value, $n_F(e)$, to the value required for reflection, $n_R(e)$, and is therefore given by

$$\tau = \frac{1 - n_R(e)/n_F(e)}{\alpha n_R(e)}. \quad \dots\dots(28)$$

Now the experiments of Bracewell and Straker (1949) have shown that τ is not more than some 150 sec.; $n_R(e)$ is about $250/\text{cm}^3$; $n_F(e)$ is unknown, but provided it is not too small the numerator of the fraction in (28) can be taken as approximately unity. Hence it would seem that α is at least of order $3 \times 10^{-5} \text{ cm}^3/\text{sec.}$ at the mean level (which may be a few kilometres below the normal height of reflection). Such a coefficient, though perhaps not impossible to accept,* is certainly large enough to cause some uneasiness. The obvious alternative to recombination is that the recovery is due to the attachment of electrons to neutral particles to form negative ions. This process could clearly proceed at a sufficient rate, but it

* For example the estimate of λ made in § 3.1 may be too low. See also footnote on page 134.

requires that during the flare the ratio of the concentration of negative ions to that of electrons should be reduced: and as far as can be judged ordinary photo-detachment would only yield a reduction if there were an appreciable increase in the *total* energy flux from the Sun—an event that seems inconceivable.

ACKNOWLEDGMENT

One of us (M.J.S.) wishes to express his thanks to the Ministry of Education for the award of a grant.

REFERENCES

- ALLEN, C. W., 1948, *Terr. Magn. Atmos. Elect.*, **53**, 433.
 APPLETON, E. V., and SAYERS, J., 1938, *Union Radio Scient. Inst.*, **78**, 272.
 BARBIER, D., 1948, *Ann. Geophys.*, **4**, 193.
 BATES, D. R., 1939, *Mon. Not. Roy. Astr. Soc.*, **100**, 25; 1947, *Terr. Magn. Atmos. Elect.*, **52**, 71; 1949 a, *Proc. Roy. Soc. A*, **196**, 562; 1949 b, *Mon. Not. Roy. Astr. Soc.*, **109**, 216.
 BATES, D. R., BUCKINGHAM, R. A., MASSEY, H. S. W., and UNWIN, J. J., 1939, *Proc. Roy. Soc. A*, **170**, 322.
 BATES, D. R., and MASSEY, H. S. W., 1946, *Proc. Roy. Soc. A*, **187**, 261; 1947, *Ibid.*, **192**, 1.
 BATES, D. R., and SEATON, M. J., 1950, *Mon. Not. Roy. Astr. Soc.* (in the press).
 BRACEWELL, R. N., and STRAKER, T. W., 1949, *Mon. Not. Roy. Astr. Soc.*, **109**, 28.
 BRADBURY, N. E., 1938, *Terr. Magn. Atmos. Elect.*, **43**, 55.
 BUDDEN, K. G., RATCLIFFE, J. A., and WILKES, M. V., 1939, *Proc. Roy. Soc. A*, **171**, 188.
 CHAPMAN, S., 1931 a, *Proc. Roy. Soc. A*, **132**, 353; 1931 b, *Proc. Phys. Soc.*, **43**, 26, 433.
 CHAPMAN, S., and PRICE, W. C., 1936, *Rep. Prog. Phys.*, **3**, 55 (London: The Physical Society).
 DUFAY, M., and DUFAY, J., 1947, *C.R. Acad. Sci., Paris*, **224**, 1834.
 DURAND, E., OBERLY, J. J., and TOUSEY, R., 1949, *Astrophys. J.*, **109**, 1.
 GOCKEL, H., 1948, *Ann. Geophys.*, **4**, 232.
 HARANG, L., 1945, *Geofys. Publ., Oslo*, **16**, No. 6.
 HOPFIELD, J. J., 1946, *Astrophys. J.*, **104**, 208.
 HOYLE, F., 1949, *Recent Advances in Solar Physics* (Cambridge: University Press).
 HOYLE, F., and BATES, D. R., 1948, *Terr. Magn. Atmos. Elect.*, **53**, 51.
 HUNTER, A., 1942-43, *Rep. Prog. Phys.*, **9**, 5 (London: The Physical Society).
 JOUAUST, R., and VASSY, E., 1941, *C.R. Acad. Sci., Paris*, **213**, 139.
 MARTYN, D. F., MUNRO, G. H., HIGGS, A. J., and WILLIAMS, S. E., 1937, *Nature, Lond.*, **140**, 603.
 MITRA, S. K., BHAR, J. N., and GHOSH, S. P., 1939, *Indian J. Phys.*, **12**, 455.
 MULLIKEN, R. S., 1942, *Rev. Mod. Phys.*, **14**, 204.
 NICOLET, M., 1945, *Mem. R. Met. Inst., Belgium.*, **19**, 1.
 PENNDORF, R., 1949, *J. Geophys. Res.*, **54**, 7.
 PRESTON, W. M., 1940, *Phys. Rev.*, **57**, 887.
 PRICE, W. C., 1942-43, *Rep. Prog. Phys.*, **9**, 10 (London: The Physical Society).
 RUDKJÖBING, A., 1940, *Publ. Kbh. Obs.*, **18**, 1.
 SCHNEIDER, E. G., 1940, *J. Opt. Soc. Amer.*, **30**, 128.
 SUTHERLAND, G. B. B. M., and CALLENDAR, G. S., 1942-43, *Rep. Prog. Phys.*, **9**, 18 (London: The Physical Society).
 U.S. Naval Research Lab. Report, 1948, No. R3120.
 VASSY, A., and VASSY, E., 1942, *Cahiers de Physique*, **9**, 28.
 WILLIAMS, S. E., 1940, *Nature, Lond.*, **145**, 68.

PHYSICAL SOCIETY CONFERENCE AT CAMBRIDGE

14th to 16th JULY 1949

Thursday, 14th July.

2.30–5.15 p.m.

Session 1.—*The regular behaviour of long and very long waves returned from the ionosphere.*

Speakers: J. A. Ratcliffe, *Cavendish Laboratory.*

K. W. Tremellen, *Marconi's Wireless Telegraph Co. Ltd.*

W. T. Sanderson, *Decca Navigator Co. Ltd.*

C. Williams, *Royal Aircraft Establishment.*

R. N. Bracewell, *Cavendish Laboratory.*

Friday, 15th July.

10 a.m.–12.30 p.m.

Session 2.—*The regular behaviour of medium and short waves.*

Speakers: W. J. G. Beynon, *University College of Swansea.*

A. F. Wilkins, *Radio Research Station, Slough.*

A. B. Whatman, *Ministry of Supply.*

2.30–4.0 p.m.

Session 3.—*The irregular behaviour associated with solar events.*

Speakers: W. R. Piggott, *Dept. of Scientific and Industrial Research.*

K. Weekes, *Cavendish Laboratory.*

R. Rivault, *Université de Poitiers*, will read a paper by
R. Bureau of the *Laboratoire National de Radioélectricité*,
Bagneux.

4.30–6.0 p.m.

Session 4.—*The formation of ionized regions.*

Speakers: K. Weekes, *Cavendish Laboratory.*

D. R. Bates, *University College, London.*

Saturday, 16th July.

10 a.m.–12.30 p.m.

Session 5.—*Irregularities in the horizontal plane in the ionosphere.*

Speakers: J. W. Findlay, *Cavendish Laboratory.*

A. C. B. Lovell, *Manchester University.*

G. Millington, *Marconi's Wireless Telegraph Co. Ltd.*

W. Ross, *Radio Research Station, Slough*

R. Rivault, *Université de Poitiers,*

ABSTRACTS OF PAPERS**SESSION 1****THE REGULAR BEHAVIOUR OF LONG AND VERY LONG WAVES RETURNED FROM THE IONOSPHERE****The Regular Behaviour of Long and Very Long Waves
Returned from the Ionosphere**

By J. A. RATCLIFFE

Cavendish Laboratory, Cambridge

These remarks will (a) define the scope of this session, (b) indicate how the main contributions fit into the plan of the discussion, and (c) outline some of the work carried out in Cambridge.

The frequencies considered lie between 10 kc/s. and 300 kc/s. Observations have mainly been near the frequencies of 16 kc/s. (GBR Rugby) and 100 kc/s. (Decca). Academic workers have generally concentrated on ionospheric reflection near vertical incidence and commercial organizations on reflection at oblique incidence. This session is concerned only with regular behaviour of the ionosphere.

By day the wave reflected at vertical incidence on 100 kc/s. is very weak (reflection coefficient $\rho=0.003$ in summer), but on 16 kc/s. is quite strong ($\rho=0.15$ in summer), and there appears to be a difference in the mechanism of reflection. Experiments made at Cambridge on other frequencies to find how this reflection coefficient changes with frequency at different times of the year will be described.

The daily variations of phase and amplitude of the vertically incident wave on frequencies of 16 kc/s. (GBR) and 113 kc/s. and 70 kc/s. (Decca) are recorded at Cambridge, and the nature of the regular variations will be described by Mr. Bracewell.

The detailed knowledge available at vertical incidence is not yet matched by such detailed knowledge at oblique incidence. Mr. Tremellen will explain what is known from commercial experience on 16 kc/s., particularly with reference to atmospheric noise, and the possibilities of very low frequencies for navigational aids. Mr. Sanderson will discuss what is known about the phase variations of waves of frequency near 100 kc/s. observed at distances up to 300 miles by the use of Decca equipment, particularly with reference to the useful range of a navigation system on that frequency. Mr. Williams will discuss what can be deduced about the ionosphere from measurements made at different distances on Decca and other navigational aids working in the same frequency band.

In thinking of navigational aids we must investigate "phase diversity" effects. Published knowledge in this field will be summarized. Since phase diversity effects are produced by irregularities in the horizontal structure of the ionosphere there is an overlap with Saturday's session, and it is proposed that only observed effects should be discussed today, whereas their bearing on ionosphere theory should be discussed on Saturday.

Two attempts are being made at Cambridge to relate the behaviour at vertical incidence on 16 kc/s. to that at oblique incidence. In one the interference pattern formed between the ground wave and the ionospheric wave has been plotted in an aeroplane to a distance of 800 km. The results will be described in outline. Mr. Bracewell will describe the other experiment, in which detailed observations of amplitude observed at distances of 200 km. and 500 km. have been compared with those observed at a distance of 90 km.

In the discussion information derived from the observation of atmospherics and the signal strengths of long wave stations will be discussed.

Very Low Frequency Propagation

By S. B. SMITH and K. W. TREMELLEN *

Marconi's Wireless Telegraph Company Ltd.

The introduction of high-frequency communications during the years 1924/28, coupled with the poor operating economics of v.l.f. traffic, diverted industrial research to the 3–30 Mc/s. band. For many years academic research was not so extensive in the v.l.f. band, and only since 1944 has interest been renewed in this subject. This revival of interest was due to the failure of high-frequency ionospheric ray propagation to provide accurate radiolocation and to the belief that v.l.f. propagation may provide a more stable and accurate navigational means.

We propose to discuss some aspects of our earlier experiments during the years 1920/1926.

The contributions will be divided into the following sections :

- (a) v.l.f. propagation, with particular reference to apparent abnormalities and other little known phenomena.
- (b) World atmospheric noise centres, seasonal and diurnal variations in direction and intensity.
- (c) Direction finding using different techniques, with special reference to reception along various geomagnetic paths.

* To be submitted to the Physical Society for publication.

The Effects of Sky-Wave on the Planning of Navigational Aids using Frequencies in the 70–130 kc/s. Band

By W. T. SANDERSON

Decca Navigator Company, Limited

The Decca system employs four stations which radiate continuously an unmodulated signal of about 1 kw. on frequencies 85.00, 127.5, 113.3 and 70.83 kc/s.—that is, in the ratio 6, 9, 8 and 5, which are phase-locked to a high degree of stability.

In the receiver the 3rd, 4th and 5th harmonics of the 85-kc/s. transmission are compared with the 2nd, 3rd and 6th of the other transmissions. The accuracy of the system depends on the stability of these patterns, and a number of observations has been made to determine the phase variation which can be expected at various ranges and times. Unlike the vertical incidence experiments, there is no absolute standard of phase against which to measure the variance of any transmission, but it is desirable to assess this factor so that the observed results at several points can be applied to find the expected pattern variance at any other point.

The results of some of the tests will be described. It appears that in English latitudes the night sky wave effects persist throughout the day during mid-winter on 70 kc/s., falling to about half the night value at 120 kc/s. During the summer daylight (April–September) the effects are so small up to 300 miles that they are difficult to measure, the standard error being less than a hundredth of a cycle.

Hawker has found that the secant of the sun's zenithal distance at noon gives a good approximation to the relative amplitude of the daylight errors throughout the year.

The implications of these results on the planning of medium range c.w. navigational aids will be outlined.

The Characteristics of Low-Frequency Radio Waves Reflected from the Ionosphere, with particular reference to Radio Aids to Navigation

By C. WILLIAMS

Royal Aircraft Establishment

The study of the performance of low-frequency radio navigation systems at great distances has provided useful data about the ionosphere. Using the Consol, Decca and POPI systems, records have been made of the changes in amplitude and phase which

result from the reflection of the signals by the E layer of the ionosphere. The data are obtained at fixed monitor receiving points, and it is found that the phase variations which are produced by the continuously changing physical state of the ionosphere are a function jointly of the base-line distance between the transmitting stations and the distance of the receiver.

Other variations of phase depend on the difference between the ground- and ionospheric-wave paths and are most readily assessed from data collected with mobile receivers. Data collected with fixed receivers and with receivers in aircraft show the existence and the effect of both types of variations. From the airborne data it has been possible to establish the relative amplitude of the ground and ionospheric reflected wave as a function of distance from the transmitters and to estimate the height of the reflecting layer and the reflection coefficient at oblique incidence during day and night. Day values of 0.05 and night values of 0.23 have been obtained at frequencies near 100 kc/s.

Measurements on Long and Very Long Waves

By R. N. BRACEWELL

Cavendish Laboratory, Cambridge

This contribution describes the regular daily and seasonal variations of waves of 16 and 100 kc/s. observed after reflection from the ionosphere at distances of 90, 200 and 500 km.

A steeply incident wave from GBR Rugby (16 kc/s.), when received 90 km. away at Cambridge, shows that the reflection coefficient at night in all seasons is about 0.5. In winter this strong reflection persists throughout the day, but at sunrise in summer the reflection coefficient falls rapidly to a steady value of about 0.15. The equivalent height of reflection fluctuates about 90 km. at night in all seasons. From shortly before sunrise until noon the height h decreases. The total decrease is 18 km. in summer and 12 km. in winter, and except during twilight closely follows the expression $h = a \log \cos \chi + \text{const.}$, where χ is the sun's zenith distance and a is about 6 km. The daily decrease in height causes an average change of 500 degrees in the phase of the sky-wave. The polarization of the down-coming wave is left-handed and approximately circular throughout the day.

Observations of the less steeply incident reflections at a distance of 200 km. differ from the above only to the extent that in winter the twice-reflected wave becomes comparable with the once-reflected.

On the Decca frequencies (113 and 70 kc/s.) the reflection coefficient by day is of a markedly lower order of magnitude. During the night the reflection coefficient is 0.25 in winter and 0.15 in summer. About an hour after sunrise in summer the reflection coefficient begins to diminish and rapidly falls to 0.003, after which the signal is lost in the noise. In winter 1948 the reflection coefficient had the relatively high value of 0.1 during the day. In the two previous winters it was no more than 0.02.

The height of reflection begins to decrease at sunrise, and in the course of the three hours during which it can be measured falls about 7 km. The polarization remains constant over the sunrise period and is probably circular. The transition from summer to winter conditions occurred in November 1948 within the space of a week, but the return to summer conditions was much slower.

When GBR Rugby is observed at 500 km. distance, multiple reflections are important in determining the behaviour of the sky wave, and the reflection coefficient appears to be rather greater at this glancing incidence. There is a departure from the strict symmetry about noon which is observed at steeper incidence, especially in summer, when the afternoon recovery appears to lag. The sunrise change is more rapid, in fact its onset is often sharply defined within a minute or two. This change occurs well before sunrise at ground level and also well before the first changes which are noticed on steeper incidence. The time of its occurrence does not depend on the solar cycle, and can be predicted, for a given date, within about 10 minutes.

SESSION 2

THE REGULAR BEHAVIOUR OF MEDIUM AND SHORT WAVES

The Application of Ionospheric Data to Short-Wave Transmission Problems

By W. J. G. BEYNON

University College, Swansea

A short survey is presented of the fundamental theory underlying the application of normal incidence ionospheric data to short-wave communication problems, with particular reference to calculating the maximum usable frequency (M.U.F.). Some aspects of the problem of applying normal incidence data on ionospheric absorption to the calculation of field strength in long distance transmission are also discussed.

A. F. WILKINS

Radio Research Station, Slough

No Title or Abstract received

(P', f) Records at Spitsbergen

By A. B. WHATMAN

Ministry of Supply

A selection of 60 (P', f) records will be shown which were made with Admiralty Type 249 equipment in Spitsbergen in 1942-43. These illustrate all the interesting normal and abnormal effects met with, and supplement those in a recent paper published in *Proc. Phys. Soc. B*, 1949, **62**, 307.

48136

~ ~ ~

SESSION 3

**THE IRREGULAR BEHAVIOUR ASSOCIATED WITH
SOLAR EVENTS****Irregular Behaviour of the Ionosphere Associated with Solar Events**

By W. R. PIGGOTT

Department of Scientific and Industrial Research

The phenomena occurring in the lower regions of the ionosphere which are associated with solar flares are briefly discussed. The types of experimental data obtained are summarized, and the geophysical significance of the results indicated.

The association of magnetic disturbances with solar flares and the main features of the corresponding ionospheric disturbance are described. It is shown that ionospheric and magnetic disturbance is not, in general, synchronous outside a zone of greatest storm activity. Two main types of ionospheric disturbance are identified:

- (i) the quasi-auroral type of disturbance—type A;
- (ii) the regular type of disturbance—type R.

A type A disturbance often includes severe disturbances in D and E regions (no reflection condition and auroral E), but its characteristic effects are in F region, where it produces an extremely rapid decrease in the ionization density, changes in height, followed by irregular variations and, if the layer is illuminated with light from the sun, a rapid recovery to normal or supernormal densities.

At night the A disturbance may be identified by the rapid recovery of the height towards normal values.

The fluctuations in the F region are closely associated with the fluctuations in the local magnetic field at the observing stations.

The type R disturbance consists of two phases: (i) a period during which the ionization density is above normal—the positive phase; (ii) a period during which the ionization density is below normal and shows a slow recovery superimposed on a regular diurnal variation—the negative phase. No connection between the magnetic variations and the R disturbance has been detected.

The relaxation time for type A is usually between a few minutes and one or two hours; for type R many hours to several days.

For a simple storm the disturbance varies fairly regularly with latitude in a given longitude zone, changing from type A with no type R in the active auroral zone, to type R with a little superposed type A, and then type R only. In the R zone the length of the positive phase increases and the depth of the negative decreases with increase of distance from the A zone until there may be no negative phase and, occasionally, no positive phase left. These disturbances are essentially unipolar and may reach to or beyond the equator. They are also most severe between a limited range of longitudes.

Some of the largest storms appear to consist of several disturbances, commencing at different times in different longitude or hemisphere zones.

The implications of these results are briefly considered.

Some Work at Cambridge on Radio Fade-outs

By K. WEEKES

Cavendish Laboratory, Cambridge

1. *Effects observed at vertical incidence on 2.0 and 2.4 Mc/s.*

Using a special technique, the changes of phase path and of amplitude of the pulse reflected from E region have been studied. Every fade-out is accompanied by simultaneous changes of the phase path which decreases 2–4 km. at the onset of the fade-out and increases again as the extra absorption decreases. The changes are different on the two frequencies used and the difference provides information about the distribution of ionization during the fade-out.

2. *Effects observed near vertical incidence on 70 and 113 kc/s.*

Using the c.w. transmissions of Decca, the sky wave has been isolated and its changes of phase and amplitude measured. The effects during a fade-out are very similar to those on higher frequencies. The phase path is reduced by an amount which may be as much as 20 km. and the amplitude decreases by a factor of at least 200 at the start of the fade-out. The phase and amplitude recover relatively rapidly as the intensity of the fade-out decreases.

3. *Effects observed near vertical incidence on 16 kc/s. and 40 kc/s.*

The outstanding fact in the study of very low frequencies is the large amplitude of the sky wave. During a fade-out the amplitude changes only slightly, and it is possible to study the phase changes throughout the fade-out. The phase path is reduced by an amount which may be as much as 20 km. and the amplitude changes by a factor of the order of 3. The phase changes appear to follow the solar flare intensity closely and to recover to the normal value rapidly when the fade-out is over. The effect on 40 kc/s. is similar to that on shorter waves.

4. An attempt will be made to outline an ionization distribution during the fade-out which would account for these observed results.

Les renforcements brusques des ondes très longues *

Par R. BUREAU

Laboratoire National de Radioélectricité

Read by R. RIVAULT, Université de Poitiers

Les renforcements brusques des ondes très longues ont été mis en évidence par l'enregistrement des atmosphériques; ils sont l'un des aspects des phénomènes radio-électriques consécutifs aux éruptions chromosphériques du soleil et sont ainsi d'un grand secours pour l'exploration indirecte de l'ionosphère. Grâce à eux, il a pu être dressé a posteriori une liste des perturbations ionosphériques à début brusque (P.I.D.B.) qui remonte à 1928, et qui met en évidence une influence très profonde du cycle solaire undécennal. Employé conjointement avec l'enregistrement du champ d'émetteurs sur ondes décamétriques, celui des atmosphériques met en lumière, dans la plupart des cas, le synchronisme des renforcements sur ondes très longues et des évanouissements brusques des ondes décamétriques. Sur les ondes supérieures à 16,000 mètres, les renforcements brusques s'amoindrissent ou disparaissent, ou se renversent même parfois.

Dans un certain nombre de cas, on peut associer une P.I.D.B. (due à un rayonnement solaire ultra-violet) au déclenchement ultérieur d'un orage magnétique (dû à un rayonnement corpusculaire) et à des troubles très profonds des radiocommunications; le délai qui s'écoule entre les deux phénomènes est en général voisin de 36 heures. Les P.I.D.B. associées aux observations d'éruptions chromosphériques et aux crochets du champ magnétique terrestre, deviennent un auxiliaire précieux dans les prévisions des orages ionosphériques.

Les études se poursuivent, de nouveaux domaines sont prospectés. Signalons des influences très rares, mais indubitables, du rayonnement cosmique; l'influence du seuil de sensibilité des enregistreurs. L'étude des anomalies de phases, entreprise à Cambridge sur les ondes très longues, complète heureusement la documentation relative aux P.I.D.B.

* This issue, p. 122.

SESSION 4

THE FORMATION OF IONIZED REGIONS

The Formation of the Ionized Regions

By K. WEEKES

Cavendish Laboratory, Cambridge

The variations of ionization density and height of maximum ionization for the E and F₁ regions agree well with the behaviour predicted by Chapman's simple theory of absorption of ultra-violet light in an isothermal atmosphere, provided it is assumed

that the electrons are lost by a recombination process the rate of which is independent of the gas pressure. The value of the recombination coefficient deduced from the observations is surprisingly high compared with the theoretical estimates.

A closer comparison of the Chapman theory with the observations reveals small but important discrepancies, and the various assumptions underlying the theory are re-examined in an attempt to discover the cause of these discrepancies. In particular it is found that the assumption that the recombination rate is independent of the pressure is not essential if the layer is formed in a part of the atmosphere containing two gases which have different vertical distributions of partial pressure. Such a condition occurs near 100 km. when the oxygen is in transition from the molecular to the atomic state.

The gases which may be responsible for the production of the various layers are briefly discussed.

Theoretical Considerations Regarding the Formation of the Ionized Layers *

By D. R. BATES and M. J. SEATON

University College, London. (Read by D. R. BATES)

The detailed mechanisms involved in the formation of the E, F₁, F₂ and D layers by solar ultra-violet radiation are discussed. Use is made of the results of some recent calculations on the continuous absorption cross section of atomic oxygen and nitrogen, and of the evidence on the ionic composition of the layers that is provided by the analysis of the emission spectrum of the upper atmosphere during twilight. Certain difficulties are emphasized and attention is directed to measurements that would be helpful in elucidating the position.

* This issue, p. 129.

SESSION 5

IRREGULARITIES IN THE HORIZONTAL PLANE IN THE IONOSPHERE

Irregularities in the Horizontal Plane in Region E of the Ionosphere

By J. W. FINDLAY

Cavendish Laboratory, Cambridge

The fading of pulse signals returned from region E at vertical incidence occurs even when care is taken to receive only one of the two magneto-ionic components in the down-coming pulse. Experiments will be described by which the variations of the phase and amplitude of signals on a frequency of 2.4 Mc/s. have been measured. The rapidity of the fluctuations in phase of the signals can be assumed to be a measure of the irregularity of the reflecting region, and the variations of this irregularity with time of day and with season will be shown.

Experiments on the amplitude of pulse signals returned to three receiving points spaced about 100 metres apart on the ground have shown that the variations of amplitude may be explained by assuming that the irregularities in region E have at any time random motions among themselves and also an overall horizontal drift velocity. The theories which allow the velocities of these random motions and of the drift motion to be determined will be outlined. The results obtained from the observations at three receiving points will be interpreted by these theories, and the values derived from the observations for the r.m.s. velocity of the random motion and for the uniform drift velocity will be quoted.

Meteor Ionization in the Upper Atmosphere

By A. C. B. LOVELL

Manchester University

Recent progress in the study of the scattering of radio waves from meteor trails makes it possible to assess within reasonable limits the total contribution of meteor ionization to the upper atmosphere and the irregularities thereby introduced. Significant changes from previous estimates, such as made by Skellett or Pierce, arise from two facts : (1) the discovery of the great daytime meteor radiants active from May to September; (2) the appreciation that only about 10^{-6} of the energy of the meteor is spent in ionization, and not the major part as previously assumed.

Observations on fairly high frequencies (e.g. 72 Mc/s.) show that the meteor trails are critically aspect sensitive at the instant of formation, but that the subsequent diffusion of the electron column gives rise to very complex types of radio wave reflections. On lower frequencies (e.g. 30 Mc/s.) a very much greater number of reflections is observed than would be anticipated for the scattering processes operative on the higher frequencies. These have been shown recently to be due to the loss of aspect sensitivity on the lower frequencies and the possible breaking up of the original column into separate clouds of ionization, thus producing multiple echoes instead of the single echo observed on the higher frequencies.

At least for frequencies above 25 Mc/s. it is thus established beyond doubt that all the transient echoes in the altitude range 80 to 120 km. are due to meteoric ionization, although there remains some doubt as to whether some of the fainter meteors may have an origin outside the solar system.

Scattering of Radio Waves from Region E

By G. MILLINGTON

Marconi's Wireless Telegraph Company

The existence of ionospheric irregularities was suggested over 20 years ago by T. L. Eckersley to explain the signals from high frequency transmitters received within the skip zone. For many years he carried out an extensive research into the scattering properties of the E layer.

The sporadic echoes scattered directly back from the E layer, unlike echoes observed at higher frequencies, show no definite correlation with visible meteors. These echoes, it is suggested, are only partly caused by meteors, and there must be another cause either in the sun or in the stars of the galaxy.

Long-distance scattering, in which the sources are illuminated, and the scattered signals received by waves reflected from the F layer, may be due to scattering centres in the E layer or on the ground. The experimental evidence must be critically examined to decide which is occurring.

Long-distance scattering can possibly cause radio reception over a route which is away from the great circle path when this is an impossible path owing to absorption or electron limitation. In the problem of frequency sharing in the H.F. band a scattering signal may be received from an unwanted station and cause interference with the wanted signal from another transmitter on the same frequency. The knowledge needed to assess likelihood of this type of interference will be discussed.

The Variations in Direction of Arrival of High-Frequency Radio Waves

By W. ROSS

Radio Research Station, Slough

The variations in bearings observed on high-frequency radio transmissions which are discussed are those due to processes occurring in the propagation of the waves. The direction-finders which have been used are free from variable instrumental errors, such as polarization errors, which in early work tended to obscure the true bearing deviations.

High-frequency bearings are subject to continuous fluctuations. When rapid fluctuations are minimized by averaging over a few minutes, and when the "wide" bearings observed at a trough of a fading cycle are ignored, there remain slow fluctuations of periods from a few minutes up to about half an hour. These fluctuations have an amplitude comparable with the rapid fluctuations. They are similar in bearings taken over the same path on neighbouring frequencies and also in bearings taken at stations up to 10 km. apart.

These fluctuations are apparently due to a fairly localized tilting of the ionosphere at angles of a few degrees over distances of some tens of kilometres.

The rapid fluctuations, which are probably due to wave interference effects, are observed even when pulses are used, though they are smaller than with c.w. signals. This suggests that there are irregularities of the ionosphere of smaller scale than those responsible for the slow fluctuations.

At certain times, for example at sunrise, systematic ionospheric tilts are observed. It is suggested that the random irregularities may be due to the spreading of disturbances produced by perturbations in the regular diurnal changes in the ionosphere. (See *Nature, Lond.*, 1947, 159, 132.)

Diffusion des échos su voisinage des fréquences critiques de F_2 *

Par R. RIVAUT

Université de Poitiers

On déduit de la théorie magnéto-ionique l'existence de trois composantes réfléchies par F_2 , de fréquences critiques f_0 , f_{x1} et f_{x2} . Les sondages à la verticale effectués à Poitiers comportent habituellement les composantes f_0 et f_{x1} , espacées de $f_H/2 = 0,6$ Mc/s. La composante f_{x2} , telle que $(f_{x1} - f_{x2}) = f_H = 1,2$ Mc/s., n'est enregistrée que très rarement. Mais l'on observe fréquemment des composantes de fréquences critiques comprises entre f_{x2} et f_{x1} , échelonnées suivant des valeurs sous multiples de f_H , 0,3 et 0,15 Mc/s. par exemple. Le plus souvent, le nombre de ces composantes est tel que l'intervalle $f_{x1} - f_0$ est entièrement occupé par des réflexions diffuses. On peut admettre que c'est la même cause physique qui produit la multiplication de ces composantes et la diffusion.

L'analyse des enregistrements présentant des réflexions diffuses montre que ce phénomène ne se produit nettement qu'entre le coucher et le lever solaires, d'autant plus fréquemment et d'autant plus tôt dans la nuit que celle-ci est plus longue. Cette diffusion cesse au moment du lever du soleil au sol, un peu après le début de l'accroissement matinal des fréquences critiques, quand la stratification diurne reparait après le désordre thermique nocturne; elle n'est pas liée aux périodes magnétiquement perturbées ni à celles de grande activité solaire.

On rencontre une autre espèce de réflexions diffuses, enregistrées à des hauteurs virtuelles de 800 à 1200 kilomètres environ, sur des fréquences immédiatement supérieures aux fréquences critiques de F_2 . Ces réflexions peuvent être intenses dans la deuxième partie de la nuit, et leur aspect net, principalement entre 3 et 5,2 Mc/s., fait penser à l'existence d'une région ionisée G supérieure à F_2 . Elles cessent généralement après le lever du soleil, mais coïncident souvent avec les périodes d'agitation magnétique et, surtout, avec les passages des principaux essaims de météorites.

* This issue, p. 126.

REVIEWS OF BOOKS

Fundamentals of Discharge Tube Circuits, by V. J. FRANCIS. Pp. x+133. 1st Edition. (London: Methuen & Co. Ltd., 1948). 6s. 6d.

There are very few books available on the fundamentals of discharge tube circuits. It is therefore very pleasing to the reviewer to see the present book which gives a survey of the whole subject.

No attempt is made to describe all the different circuits in common use, rather the author has attempted to discuss the more important fundamental properties of discharge tubes in relation to the circuits in which they are, or could be, operated. Chapters are included on the discharge tube as a circuit element, operation on D.C. supplies, dynamic characteristics, voltage and current wave-forms on A.C. supplies, operation on alternating current supplies, initiation and typical discharge lamp circuits and general principles of design.

The book should prove interesting and useful to a wide circle of readers particularly those who are actively concerned with the use and applications of discharge lamps. It is an authoritative account and can be recommended. D.T.

Optical Design and Lens Computation, by B. K. JOHNSON. Pp. viii+175. 1st Edition. (London: The Hatton Press Ltd., 1948.) 36s.

A former Lord Rayleigh once complained that every writer on Optics used a different notation and none of them read what the others had written. There is no doubt that this variety of notation has been the cause of the absence hitherto of any book dealing in a practical manner with the design of lens systems. Each professional lens designer has used his own notation and has been deterred from writing papers or books on the subject by the impossibility of using a universal language. Thus the old Optical Society died from want of matter to publish.

The late Professor Conrady, for the purpose of his lectures, employed a notation all his own, the attraction of which was the facility with which it could be reproduced on the typewriter and the duplicating machine, awkward as it was in other respects. His book, *Applied Optics and Optical Design*, introduced his notation and formulae to a wider circle than that of his students and a facsimile edition published in the U.S.A. during the late war carried it across the Atlantic. The Conrady notation and formulae are therefore known and can be easily referred to. Publication of his work was never completed; his book is limited to telescope objectives and eyepieces. However, his lecture work continued on to photographic lenses and microscope objectives.

The present author makes use of the Conrady notation and deals in succession with the design of telescope objectives, eyepieces, photographic lenses and microscope objectives, reproducing the treatment Conrady gave in his lectures. The author pays generous tribute to Conrady's work.

By the design of optical systems should be implied a method of working out the forms of the constituents by anyone inexperienced in practice, that is, without a background of experience to start from. For this one needs an algebraic set of equations based on the aberrations, which may be deduced from the purely geometrical considerations of ray paths, by employing the principle of least time, or by consideration of the wave surface. The author uses none of these—his method is based on ray tracing. It is doubtful if the methods he displays are the best for training opticians, but they are presumably those taught in the Technical Optics Section of the Imperial College of Science and Technology, the only school of optical designing in the country.

The methods developed in this book will not appeal strongly to the mathematician, which is a pity, since there is a very great need to attract able mathematicians to the optical trade. The methods are directed rather to the student of limited mathematical attainments. The difficulty seems to be that mathematicians of originality are repelled by the drudgery of routine calculation. Those who are not repelled often lack originality.

The present book will have value therefore for students and may interest others subject and may even stimulate other designers to publish their techniques ; some of may prove more attractive to mathematicians.

The book is well provided with examples : these are all worked out logarithmically are of course ray tracings. The clumsiness of design by ray tracing is well illustrated by chapters on eyepieces and photographic lenses, where many pages are taken up by a treatment that could be dealt with algebraically in few. It is curious that the author deals exhaustively with the coma of eyepieces but is uninterested in field flatness, which is usually more important. Likewise, in his treatment of the photographic meniscus he is at pains to show how coma can be cured by means of the stop, while in practice the coma is used to cure the astigmatism. Some of the statements made are of doubtful validity and some of the suggestions of doubtful practical value : e.g. that flint glass should be tried for eyepiece lenses. It may be argued that this is an exercise for the student, but in a practical book it should be pointed out that this is not good practice, and why.

It is regrettable that the book is written in such bad English ; "alternately" for "alternatively" on p. 71 may be a printer's error, but such constructions as the following (chosen at random) abound : "Having determined the astigmatism . . . the true curvature of field must now be determined. In order to do this, it is necessary . . ." Unfortunately there is much scientific writing in this style.

The author is to be congratulated for his pioneering work in a difficult field. H. W. L.

Five-Figure Tables of Mathematical Functions, by J. B. DALE. 2nd Edition.

Pp. viii + 121. (London : Edward Arnold & Co., 1949). Price 6s.

The first edition of this book must be familiar to most readers of this review. Since 1903 it has been one of the few convenient single-volume sets of miscellaneous tables compiled with the physicist and applied mathematician in mind. In addition to logs, antilogs and trigonometrical functions, it contained Bessel functions, gamma functions, elliptic integrals and Legendre functions, which are all of value at times to the ordinary physicist.

In the present edition the contents have hardly been altered; the table of logarithms is somewhat improved and there is now a table of the logarithms of hyperbolic functions. The old edition gave logs for the range 1.00(0.01)9.99, and the new gives, in addition, 1.000(0.001)2.999, which is a definite improvement. In the old edition the amount to be added for the next figure of the argument was given in the way that is usual with four-figure logarithms—columns for 1, 2, . . . 9, based on the mean differences of the line, were given opposite the entries. In the new edition this plan is retained for the first table, but for the second the actual differences are to be used, and there is a proper table of proportional parts, but, unfortunately, it is not on the same page as the logs themselves.

The whole book has been re-set, with modern "heads and tails" type, instead of the old, uniform-height type, and on paper which is very pleasant to handle. J. H. A.

Electronics in the Factory, edited by H. F. TREWMAN. Pp. 188. 1st Edition. (London : Pitman, 1949.) 20s.

This book is evidently intended to keep informed the works director or general manager—in short, the executive—as to the more general applications of electronics in industry. Examples of electronic devices for timing, counting, measurement, process control, eddy-current heating, etc., are discussed in simple terms which make a reasoned appeal to the imagination.

To the professional physicist the book makes light reading; I disposed of it in a restful evening and enjoyed it. G. I. F.

A Text Book of Heat, by LEROY D. WELD. 1st Edition. Pp. x+436. (New York : The Macmillan Co., 1948). Price 25s.

This is described as a book for upper classmen, and has been based on the material used for 35 years as lecture notes in a junior-senior course in heat. This does not mean that it is old-fashioned, for modern work has been incorporated wherever it is needed. The properties of liquid helium receive due attention, the totally enclosed absorption type of refrigerator is described, and the International Temperature Scale properly treated.

The general outlook of the book is good. It aims evidently at presenting the student with principles rather than with practice, and many teachers may feel that it gives too little experimental detail; whether he is telling the student how to measure heat quantities or temperatures, or describing the evidence which led to the acceptance of the first law of thermodynamics, the author gives clear explanations of the principles of the experiment, but very little about the practical dodges adopted or required.

The scope of the book includes thermodynamics and statistical mechanics, but the latter suffers because quantum theory is always treated as something outside, to be superimposed on the classical theory. Apart from this, it is an excellent treatment.

The chapters are all accompanied with well selected sets of exercises, and with references for further reading. These show a distinctly American bias, which is, indeed, evident throughout the book. The attribution of the theory of light darts to Epstein may be an example of this, or may be a misprint.

As a whole, the book is very sound, and may give some very useful ideas and hints to those engaged in teaching this subject at the intermediate or pass degree level. J. H. A.

Theory of Dielectrics : Dielectric Constant and Dielectric Loss, by H. FRÖHLICH. Pp. vi+180. 1st Edition. (Oxford : Clarendon Press, 1949.) 18s.

The field of Professor Fröhlich's book is fairly precisely defined in its subtitle, "Dielectric Constant and Dielectric Loss": the contents being confined to theories of dielectric substances situated in ordinary (weak) fields for which there is no question of dielectric saturation or breakdown. The treatment, moreover, is entirely classical and there is no discussion of specifically quantal effects significant only at low temperatures (or for very strong fields). Within these limits, however, it deals with its subjects thoroughly.

There are four chapters. The first briefly defines the macroscopic concepts with which we are concerned; in particular, the ordinary static dielectric constant ϵ and the complex dielectric constant, $\epsilon_1 + i\epsilon_2$, necessary in dealing with alternating fields. The quantities ϵ_1 and ϵ_2 are shown not to be independent functions of frequency; and the energy loss is proportional to ϵ_2 .

Chapters II and III between them comprise over half the book, and are concerned with theories appropriate to the macroscopically static and dynamical cases, respectively. The former chapter deals with the Clausius-Mosotti formula, Onsager's extension of this, Kirkwood's theory of polar liquids and some rather general, and elegant, theorems due to the author. The applicability of these formulæ, to solids, liquids, gases, or solutions, is discussed in terms of the physical assumptions made: indeed, in the reviewer's opinion, one of the most useful features of the book is its classification of a wide range of dielectric problems in terms of the physical forces of primary importance. The latter chapter, concerned with relaxation effects, again classifies these according to their mechanisms, in particular distinguishing between loss due to elastically bound charges (resonance absorption) and that due to transitions of charges, or dipoles, over potential barriers. In the second case several models that lead to exponential decay functions are discussed in some detail, as also are the ensuing Debye equations. The effect on the frequency dependence of energy-loss of a spread of relaxation times is also investigated.

The final chapter deals with applications. In contrast to the previous chapters which had been heavy going, the reviewer found this rapid survey of a considerable range of experimental data very delightful reading. No attempt is made to include all existing data relevant to the theme of the book, but certainly the ground is well covered, and the chapter discusses such diverse topics as the dielectric constant of water (Kirkwood), the torsional

flexibility of ketone molecules in paraffin solutions, and the ammonia absorption spectrum for centimetre waves (work of Bleaney and Penrose). This chapter, particularly, raises further problems for both theoretical and experimental work.

It ends, unfortunately, with a section, chiefly theoretical, on ionic crystals which is almost too condensed to be intelligible. Indeed the reviewer feels that perhaps the theoretical sections are, as a whole, rather over condensed and that sometimes greater clarity in the writing might have been achieved. It is possible that a more readable account would have resulted from a parallel development of the theory and the experimental material rather than a consecutive one. Despite this and a few minor defects (such as the omission of experimental points from Figure 19 and that the assumption that ϵ is independent of the field strength limits the general validity of some of the thermodynamical considerations) the book remains a very useful and stimulating addition to the literature.

The book is well produced, as one expects from the Oxford University Press ; and there is a useful list of nearly a hundred references to original papers. G. S. R.

CONTENTS FOR SECTION A

	PAGE
Dr. H. FRÖHLICH and Mr. J. O'DWYER. Time Dependence of Electronic Processes in Dielectrics	81
Mr. J. H. SIMPSON. The Time Delay in Conduction and Breakdown Processes in Amorphous Solids	86
Dr. E. BILLIG and Dr. P. T. LANDSBERG. Characteristics of Compound Barrier Layer Rectifiers	101
Dr. J. V. JELLEY and Dr. E. B. PAUL. Fast Neutron Reactions with Fluorine and Sodium	112
Mr. J. M. HILL and Mr. L. R. SHEPHERD. Slow Neutron-induced Activities in Europium and Samarium	126
Mr. H. ELLIOT and Mr. D. W. N. DOLBEAR. Directional Measurements of the Diurnal Variation of Cosmic-Ray Intensity	137
Dr. C. C. BUTLER, Mr. W. G. V. ROSSER and Mr. K. H. BARKER. Some Properties of Penetrating Cosmic-Ray Showers and Star Phenomena seen in the Cloud Chamber	145
Miss B. CHOWDHURI. Experiments on the Nature of Penetrating Events in Extensive Air Showers	165
Letters to the Editor :	
Mr. F. K. GOWARD, Dr. E. W. TITTERTON and Mr. J. J. WILKINS. Observations of (γ , T), (γ , D) and (γ , np) Reactions in Boron	172
Mr. D. A. WRIGHT and Dr. R. A. WEALE. Conductivity of Thin Films	173
Dr. A. F. GIBSON and Mr. T. S. MOSS. The Photoconductivity of Bismuth Sulphide and Bismuth Telluride	176
Mr. J. J. WILKINS. Validity of Two-dimensional Design of Synchrotron Pole-Faces	177
Mr. R. BOWERS and Dr. K. MENDELSSOHN. Properties of Superflow in Liquid Helium II	178
Reviews of Books	181
Contents for Section B	182
Abstracts for Section B	183

ABSTRACTS FOR SECTION A

Time Dependence of Electronic Processes in Dielectrics, by H. FRÖHLICH and J. O'DWYER.

ABSTRACT. In solid dielectrics electrons may be excited from low levels into conduction levels or into traps with energies slightly lower (shallow traps) by an ionizing radiation or by the action of strong fields. It is assumed that collisions amongst electrons in these levels are so frequent that an electronic temperature T is established which may be different from the lattice temperatures. Also the number of electrons in these levels need not correspond to equilibrium at temperature T , in which case transitions to or from lower levels (of lattice defects or foreign admixtures) will take place. Such transitions are either due to excitation by sufficiently fast-moving electrons in the conduction levels, or to the inverse process. In this way it is possible to account for recombination within a reasonable time provided these lower levels (denoted as deep traps) are not much more than 1 electron volt below the lowest of the shallow traps. The difficulty of having to introduce recombination with simultaneous emission of very many vibrational quanta, an extremely unlikely process, is thus avoided. Other applications of this method, for example to the time dependence of electronic conductivity in strong fields, are indicated.

The Time Delay in Conduction and Breakdown Processes in Amorphous Solids, by J. H. SIMPSON.

ABSTRACT. A method of estimating the time delay in changes of conductivity and in the breakdown process for amorphous materials, based on H. Fröhlich's high-temperature breakdown theory, is given. Numerical values of these times are obtained for the case of soda-lime glass, and it is estimated that the time constant of the conduction process is between 10^{-4} and 10^{-5} second, while that of the breakdown process may be considerably shorter. Experimental data obtained by Turner and Lewis indicate that the estimate for the conductivity process is of the correct order.

Characteristics of Compound Barrier Layer Rectifiers, by E. BILLIG and P. T. LANDSBERG.

ABSTRACT. The assumptions involved in Mott's and Schottky's theories of rectification are analysed and a potential barrier at the metal-semiconductor interface is proposed which enables one to pass continuously from one theory to the other. The model is directly applicable to the practically important case of rectifiers possessing an additional insulating layer between semiconductor and counter-electrode. The current-voltage characteristic of such a compound barrier is obtained and contains Mott's and Schottky's results as special cases. In this connection, an expression for the volume distribution of current carriers is obtained and plotted for the case of a linear potential distribution in the barrier. A discussion of the distribution of charges throughout the rectifier leads naturally to the concept of a capacitance of the barrier, and to the derivation of the capacitance-voltage characteristic. In this connection, a definite experimental problem is suggested by the analysis.

Fast Neutron Reactions with Fluorine and Sodium, by J. V. JELLEY and E. B. PAUL.

ABSTRACT. The yields of the (np) and $(n\alpha)$ reactions in fluorine and sodium have been measured by observations on the radioactivity produced by bombardment with neutrons from the (dn) reactions in Li, B, Be and D. Cross sections have been estimated and methods of deriving the absolute disintegration rates and the neutron flux are described.

The reaction energies have also been derived, by assuming a particular form for the excitation function, and using published data on the energy distribution of the neutrons used.

Slow Neutron-induced Activities in Europium and Samarium, by J. M. HILL and L. R. SHEPHERD.

ABSTRACT. The activities produced by the neutron irradiation of europium and samarium have been investigated by means of a thin lens type β -ray spectrometer and by absorption and coincidence methods. The 9.2-hour activity in europium is due to the dual decay of ^{152}Eu by β^- emission and by K capture. The long period activity is due to both ^{152}Eu and ^{154}Eu , and again both β^- emission and K capture are present. This long period K capture is due to ^{152}Eu , and the β^- emission is possibly due to both ^{152}Eu and ^{154}Eu . The energies of the emitted radiations are discussed and possible decay schemes proposed.

The 47-hour activity induced in samarium is due to the β^- decay of ^{153}Sm , and measurement of the energies of the emitted radiations has been made.

Directional Measurements of the Diurnal Variation of Cosmic-Ray Intensity, by H. ELLIOT and D. W. N. DOLBEAR.

ABSTRACT. Measurements of the solar daily variation in cosmic-ray intensity have been made for a total of 360 days using two arrays of Geiger-Müller counters inclined at 45° to the vertical in the North and South directions. Both the diurnal and semi-diurnal variations are found to be significantly different for the two directions, and since the particles recorded by the two sets have passed through the same amount of atmosphere under similar conditions, it must be supposed that at least a part of the variation is due to causes which lie outside the earth's atmosphere.

It is found that both the amplitude and phase of the daily variation for the South-pointing recorder vary with season, whereas in the North direction the amplitude only shows any seasonal variation. The amplitude is greatest in summer in both directions.

The experimental results are compared with those of earlier workers and are discussed in relation to some of the interpretations of the daily variation which have been put forward. The difference between the variations in the two directions is interpreted as evidence for a non-isotropic distribution of the primary cosmic rays in space. A variation with sidereal and not with solar time would be expected under these conditions, but it is pointed out that the magnetic field of the sun may modify the direction of arrival of the primary rays at the earth so as to produce a resultant solar time variation.

Some Properties of Penetrating Cosmic-Ray Showers and Star Phenomena seen in the Cloud Chamber, by C. C. BUTLER, W. G. V. ROSSER and K. H. BARKER.

ABSTRACT. Cloud-chamber photographs of cosmic-ray showers containing penetrating particles have been obtained at sea level. The spectrum of the penetrating particles has been measured and a study made of the interaction of the energetic particles in lead and in the gas of the cloud chamber. Conclusions about the nature of the penetrating particles are discussed; these show that about half the penetrating particles are protons and the remainder are probably π -mesons and mesons of heavier mass. Three heavy particles produced in stars are probably τ -mesons.

Experiments on the Nature of Penetrating Events in Extensive Air Showers, by B. CHOWDHURI.

ABSTRACT. The discharges, in coincidence with extensive showers, of individual counters of a group placed in a cavity under 15 cm. lead have been studied with particular reference to the number of these counters simultaneously discharged in a single event. It is shown that the observed multiplicity of discharge is not consistent with operation by particles, each discharging one counter only, distributed at random in the extensive shower falling on the apparatus. The multiplicity is also inconsistent with that expected for low energy photons at the extreme range of normal cascades, and with that arising from the secondary electrons of a reasonable μ -meson spectrum. Entry of energy into the shielded cavity is shown to take place over an area much smaller than the total area of the cavity, but the mechanism then leading to the observed multiplicity has not been identified.

THE PROCEEDINGS OF THE PHYSICAL SOCIETY

Section B

VOL. 63, PART 3

1 March 1950

No. 363 B

The Deposition of Ice Crystals on Cooling Surfaces in Low Temperature Plant

By G. G. HASELDEN

Chemical Engineering Dept., Imperial College of Science and Technology, London S.W.7

MS. received 17th May 1949, and in amended form 28th August 1949; read before the Low Temperature Group 17th December 1948

ABSTRACT. A study is made of the conditions governing the deposition of ice crystals from a moist air stream on to the inside surface of a cooled tube. The promoting effect of impurities such as oil vapour in the air stream is demonstrated. Series of experiments were made to establish the effect of air velocity, and of the temperature difference between the tube wall and air stream, on the amount of deposition. In the flow range tested (Reynolds' number 11,000 to 27,000) it was observed that increasing the air velocity increased the percentage deposition; at the lower velocities there was no evidence of progressive deposition. The effect of temperature difference on the percentage adhesion in the range 7° to 14° C. was small, but an experiment at a value of 4.6° C. showed a decrease in the amount of condensed snow deposited. In most experiments a glass tube was used but in others use of a brass tube gave similar results.

§ 1. INTRODUCTION

THE presence of small concentrations of condensable impurities, such as water vapour, in gases processed in a low temperature plant will lead to blockage of the plant unless some method of purification is adopted. The method of removal may use refrigeration or it may employ a process external to the low temperature part of the plant. Where external means of purification are employed, such as drying with alumina gel, the removal is usually incomplete and some further device must be used if frequent shut-downs are to be avoided.

The refrigeration methods of purification normally employ duplicate heat exchangers or regenerators. Whilst one of the exchangers or regenerators is in use, the other is allowed to thaw out, or the condensed impurity may be revaporized by passing a return stream of gas over it. In both types of cooling apparatus it is desirable that the condensed solid should immediately adhere to the cooling surface and not be carried forward to a colder zone of the plant where it is more difficult to remove. An exception to this practice is the process described by Kornemann and Yendall (1944) in which a heat exchanger is so designed that carbon dioxide condensing from air at temperatures below about -110° C. is maintained completely in suspension and is subsequently removed by scrubbing the gas with liquid air.

Little information is available concerning the processes by which crystalline precipitates are formed and adhere to the cooling surface under the conditions prevailing in low temperature plant, but it is often apparent that all the condensed

solid does not adhere to the cooling surface and that a part is carried on in suspension in the gas stream. It was the purpose of these experiments to study the deposition process and to make a number of measurements of the percentage adhesion of ice crystals condensing from a moist air stream under various flow and temperature conditions.

§ 2. APPARATUS

The deposition of ice from a moist air stream was made to take place on the inner surface of a tube of approximately $\frac{1}{8}$ inch inside diameter and 12 inches long, the tube being cooled by circulating cold ethyl alcohol around it. The design of this experimental heat exchange tube is shown in Figure 1. In the majority

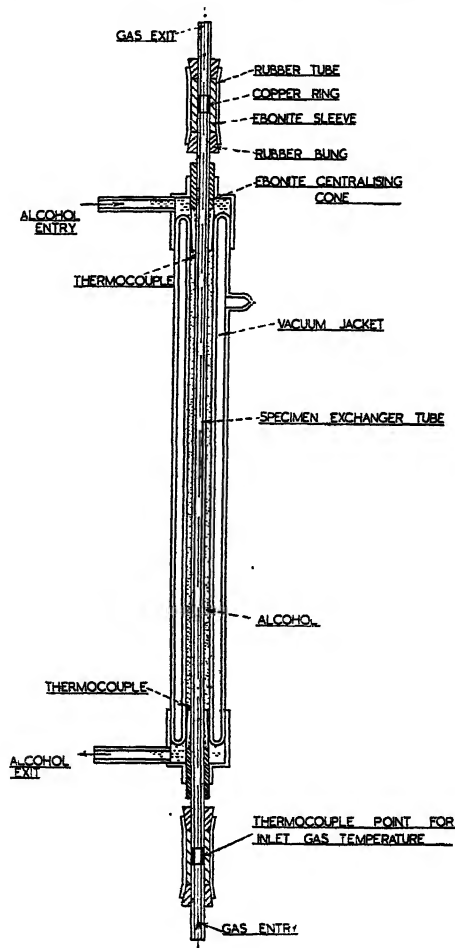


Figure 1.

of the experiments a glass tube was used, to permit visual inspection of the deposit, though a few trials were made with a brass tube. The tube was positioned at the centre of a vacuum jacketed sleeve by means of ebonite cones. The cold alcohol was brought into the upper brass header and flowed through slots in the ebonite cones into the annular space between the vacuum jacket and the experimental tube. The alcohol was drawn off continuously through the lower brass

header. The temperature of the alcohol was measured by copper-constantan thermocouples inserted in the slots in the two ebonite cones.

In measuring the temperature of the air before entering and after leaving the experimental tube it was essential not to cause any obstruction of the gas path, or building up of snow might result. The method used was to insert into the air line, at either end of the experimental tube, a copper ring of the same diameter which was held in alignment by an ebonite sleeve. The copper ring achieved very closely the temperature of the air stream and its temperature could be measured by a thermocouple soldered to it.

The precooling and method of connection to the experimental heat exchange tube is shown in Figure 2. It is important that the condition of the air entering

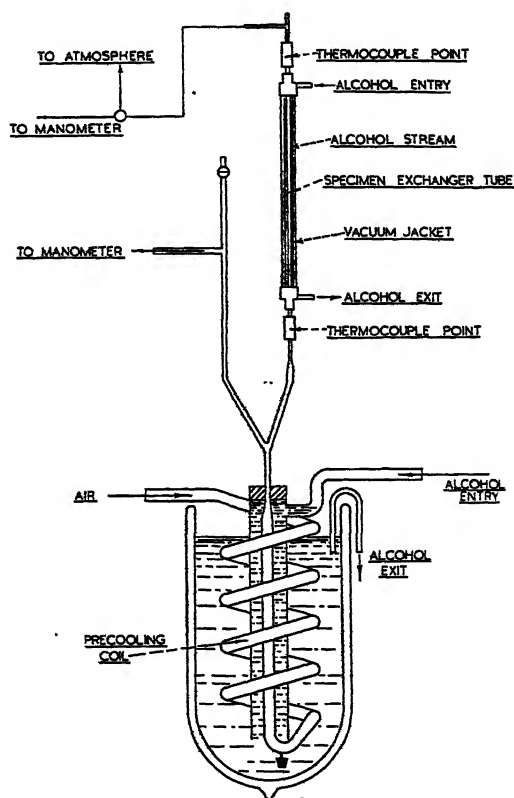


Figure 2.

the specimen tube should resemble as closely as possible that prevailing at an intermediate point in a commercial exchanger. For this reason the precooling is carried out counter-currently, the final limb of the cooler being straight so that the 'carry-over' of ice crystals produced in the final stages of precooling would not be prevented. The precooling coil was of large diameter glass tubing to accommodate the ice deposited in experiments of long duration. The tube leading from the precooling coil to the exchanger was tapered gradually in two places and was provided with a 'Y' connection at the centre. The purpose of this connection was to provide a lead to the manometer and also to allow pre-cooled air to be blown off whilst the temperatures were being stabilized.

§3. LAYOUT

The layout of the apparatus and design of the cryostat for maintaining temperatures down to -50°C . are shown in Figure 3. Separate alcohol streams were used to cool the precooler and exchanger, the temperature gradient between the two being adjusted either by varying the alcohol flow or by applying different currents to the electric heaters in each stream.

In order to continue the experiments for periods up to seven hours it was found necessary to reduce the humidity of the air entering the precooler. It was found

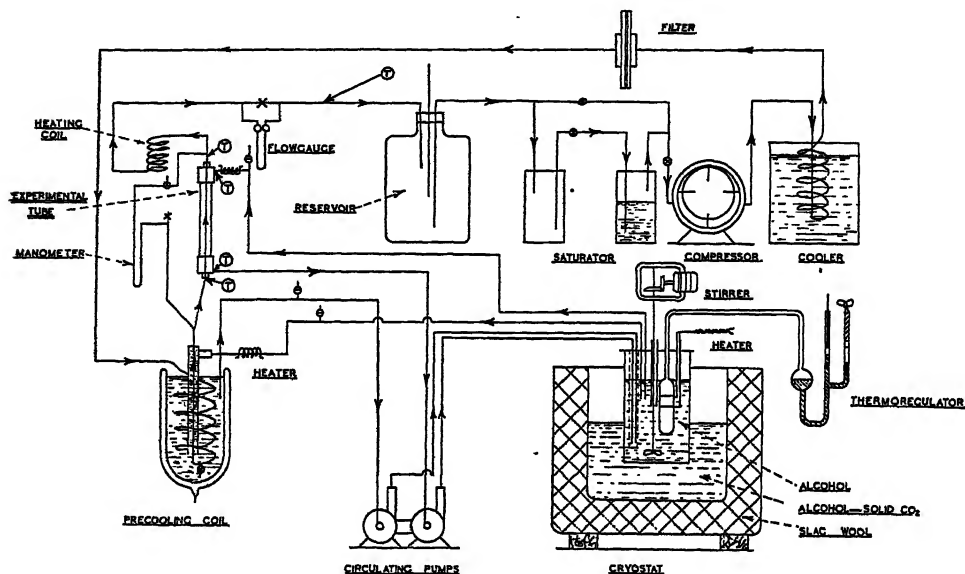


Figure 3.

more convenient to do this by continuously re-cycling the air and adding water vapour either by bubbling a fraction of the re-cycled air through water or more usually by admixing atmospheric air with re-cycled air at the compressor suction. The air leaving the compressor was water-cooled and passed through a glass-wool filter before being admitted to the precooler.

The manometer connections to the experimental tube were so arranged that either the inlet pressure to the experimental tube, or the pressure drop across it, could be measured.

§4. EXPERIMENTAL OBSERVATIONS

The first experiment was made under the following conditions: air flow rate $1.4\text{ ft}^3/\text{min.}$, mean temperature of alcohol in exchanger -40°C ., temperature of precooled air -34°C .

The temperatures of the precooler and tube were brought to the required value whilst the precooled air was being blown to waste, then the air stream was diverted to the tube. After about 10 minutes (the time taken for the connecting tube and insulation to cool down) the inside surface of the tube frosted over. This deposit was seen through a lens to be of uniform dendritic structure growing from points fairly evenly disposed over the tube surface. During the first 30 minutes the deposit was slightly denser towards the top of the tube, but this became progressively less apparent as the experiment proceeded. The particle

size appeared to be of the order of 100μ but it was impossible to see either the intricacies of the structure or the way in which the deposit was forming. After about 2 hours the deposition became less regular and narrow shelves of thicker deposit, lying at right angles to the direction of flow, were formed. The pressure drop across the tube increased from 10 cm. Hg to about 22 cm. Hg and the experiment was discontinued.

The apparatus was allowed to warm up and the condensates from the tube and the precooler were examined and were found to contain suspended oil droplets. The experiment was repeated and all the observations confirmed.

Since the air leaving the compressor was filtered through glass-wool the actual amount of oil depositing in the tube was small (less than 5% by weight), but it appeared possible that this oil was exerting a large influence on the deposition process. More efficient oil removal was therefore attempted by inserting a packed copper oxide tube, maintained at dull red heat, in the compressor delivery and filtering the effluent air, after cooling, through a paper filter. A repeat experiment showed that the rate of increase of pressure drop across the tube was reduced to approximately 20% of that experienced in the first experiment. Inspection of the condensate after thawing showed, however, that the oil removal was still incomplete, the length of the copper oxide tube probably being inadequate.

A third experiment was then made in which the rotor of the compressor was replaced by one adapted to take carbon blades. All parts of the compressor were cleaned with carbon tetrachloride and the compressor was thereafter run without oil. The third experiment again showed deposition starting after about 15 minutes, but subsequent deposition occurred very slowly and the structure of the deposit was finer than before. The increase in pressure drop in this case was only 4.0% of that occurring in the first experiment.

Even in the third experiment the condensate in the precooler was very slightly clouded but there was no sign of an oil film in the tube. From these results it is apparent that the presence of oil vapour exerts a very large influence on the deposition process.

§ 5. EFFECT OF AIR VELOCITY

Using the non-lubricated air compressor and the paper air filter a number of experiments were made for different air flow rates, the temperatures being maintained approximately constant. The experimental tube was carefully cleaned before each experiment by drawing through the tube a cotton-wool plug soaked in acetone.

The results of these experiments are shown in Figure 4. The slope of these curves indicates that increasing the gas velocity increases the amount of adhesion, but a more complete analysis of the results is necessary to establish the quantitative relationships. This is complicated by the fact that a change of flow rate not only causes a direct change of pressure drop across the tube but also changes the amount of water vapour admitted to the tube, the mean pressure of the system, and the exit air temperature.

The variable it is required to follow in any experiment is the proportion of the snow condensed which actually adheres to the tube wall. Although it is not possible to measure either of these quantities directly, it is possible to estimate them from the pressure, temperature and flow readings. If the effects of super-saturation are small the amount of snow being condensed is given by the product

of the air flow rate and the difference in vapour pressure of ice at the inlet and outlet air temperature. The amount of snow adhering to the tube can be found indirectly by calculating the reduction of area of the tube by substituting the experimentally determined pressure drop in a formula for the flow of gases in tubes. In using such a formula, correction can automatically be made for small variations of mean pressure and temperature from one experiment to another. From the percentage reduction in area of the tube the thickness of the crystal layer or volume of the deposit can be found. From an assumed value of the bulk density of snow under these conditions the weight of snow adhering to the tube may be calculated.

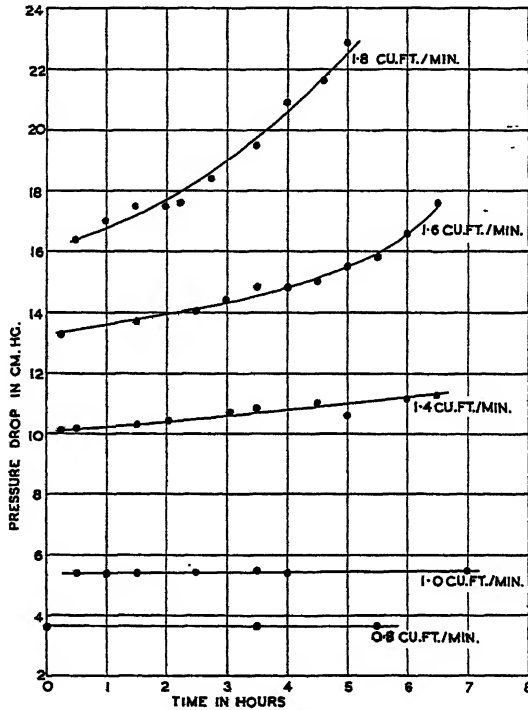


Figure 4.

The following formula for the flow of compressible fluids through pipes (McAdams 1942) was used:

$$\Delta p = \frac{G^2}{g_c} (v_2 - v_1) + \frac{2f_m G^2 L}{g_c \rho D}$$

where Δp = pressure drop across tube, G = mass flow per unit area, g_c = gravitational constant, v_1, v_2 = initial, final specific volumes, f_m = friction factor, L = length of tube, ρ = mean density of fluid, D = tube diameter = 0.129 in. (0.01075 ft.).

In applying this formula it was found that the first term was always less than 1% of the total and could therefore be neglected. Hence

$$\Delta p = \frac{2f_m G^2 L}{g_c \rho D^5}.$$

If x is the total mass flow per unit time and C_1 , C_2 are constants depending on the units employed, then

$$\Delta p = C_1 \frac{f_m x^2 L}{\rho D}.$$

For the range of Reynolds' numbers R from 5,000 to 200,000 Koo (1932) states that

$$f_m = \frac{0.046}{(R)^{0.2}} = \frac{0.046}{(4x/\pi D \mu)^{0.2}},$$

therefore by substitution,

$$\Delta p = C_1 \frac{x^2 L}{\rho D^5} \cdot \frac{0.046}{(4x/\pi D \mu)^{0.2}} = C_2 \frac{x^{1.8} L \mu^{0.2}}{D^{4.8} \rho}$$

or

$$D^{4.8} = C_2 \frac{x^{1.8} L \mu^{0.2}}{\rho \Delta p}.$$

Both x and ρ can be determined from the experimental readings and the value of C_2 can be calculated. From a test experiment in which no deposition occurred the calculated value of D was 0.01072 ft., which compared closely with the measured value of 0.01075 ft.

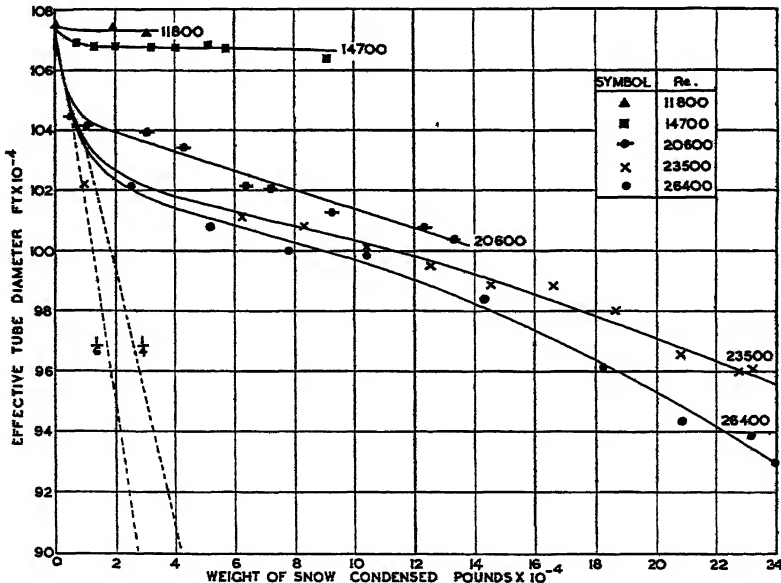


Figure 5.

In applying the above equation to the experimental results a further correction must be made for the uncooled parts of the tube, in which no deposition is assumed to occur.

Figure 5 shows the results for varying flow rates (expressed as Reynolds' number), plotted according to this revised method. The dotted curves show the decrease in diameter which would occur if all the condensed snow adhered to the tube wall, assuming a bulk density of snow of $\frac{1}{2}$ and $\frac{1}{8}$ respectively.

From this graph it appears that the initial rate of deposition was high but that it fell off rapidly to zero in the case of low gas velocities and to a value of approximately 10% at higher velocities. The first layer of ice crystals to deposit will present a roughened surface to the gas flow and therefore will produce a pressure effect out of proportion to the layer thickness. Therefore the initial slope of these curves may be somewhat steeper than is actually the case.

§ 6. EFFECT OF TEMPERATURE HEAD

A further series of experiments was made in which the air flow rate was maintained constant at 1.4 ft³/min. and the inlet air temperature at approximately -31°C., whilst the temperature of the alcohol in the heat exchanger was varied from approximately -36° to -50°C.

The results of this series are shown in Figure 6, and with the revised system

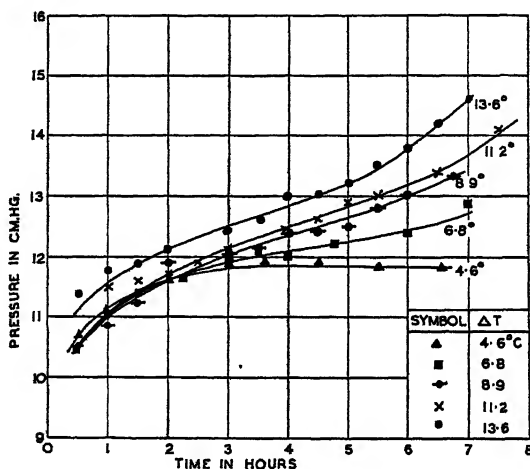


Figure 6.

of representation in Figure 7. The temperatures shown in these diagrams are the mean temperature differences between the air and alcohol and represent closely the temperature differences between the air and tube wall. From Figure 7 it is apparent that the effect of temperature head in the range tested is small except in the case of the lowest value which gives evidence of zero adhesion. No exact comparison can be made of the slopes of the curves during the early part of the run because these are very largely influenced by the starting conditions of the experiment which could not be exactly reproduced.

§ 7. EXPERIMENTS USING A BRASS TUBE

A few experiments were made using a brass tube of 0.127 in. diameter in place of the glass tube previously used. The results of these experiments were very close to those for the glass tube formerly used, thus demonstrating that the material of the tube has little effect on the adhesion process.

§ 8. DISCUSSION OF RESULTS

In assessing the validity of these results the possibility of supersaturation of the air-stream must be considered. If the air flowing through the exchanger becomes supersaturated instead of precipitating solid the results of all the experiments would be misleading. This state of affairs is considered unlikely

for the following reasons. The cooling surface of the precoolers will rapidly become coated with ice crystals and further crystals will be held in suspension in the air stream, supplying abundant nuclei on which further condensation can occur. The connecting tube between the precoolers and exchanger is short and heavily lagged so that the air entering the exchanger will still contain suspended crystals. In the exchanger the condensation will take place under similar conditions with a coating of ice crystals on the cooling surface and other nuclei present in the air stream. In the experimental tube supersaturation will be less likely, due to the higher degree of turbulence. The fact that the deposition of crystals was uniform over the whole length of the tube, except in the cooling down period when the air entering the exchanger was superheated, confirms the fact that the supersaturation is slight. If supersaturation were occurring to any great extent the deposit at the top of the tube would be much denser than the deposit at the bottom end.

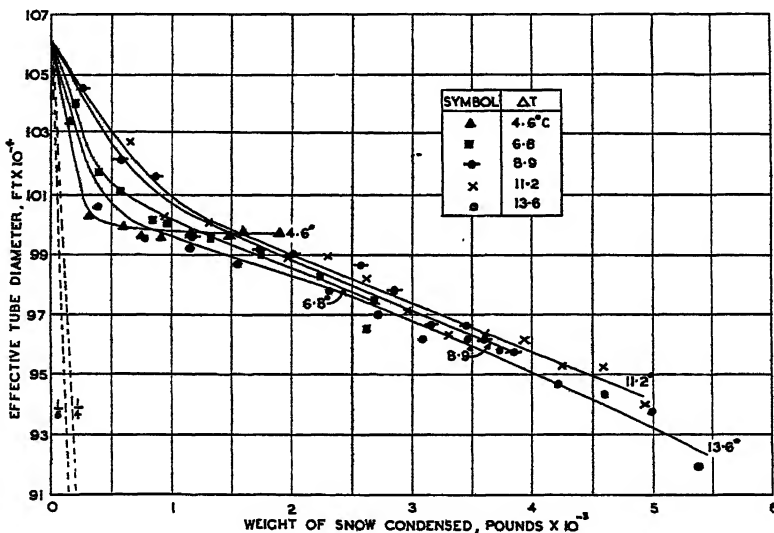


Figure 7.

In applying these results to an understanding of the processes occurring in a low temperature plant it is at once apparent that unless a considerable quantity of oil is present in the entering gas it is very unlikely that all the precipitated solids will remain in contact with the cooling surface. In the case of regenerators of the conventional design the much lower gas velocity, combined with the tortuous gas path and the very small thickness of the deposit during each change-over period, may produce a greater amount of adhesion. In the case of tubular heat exchangers it seems possible that over a certain temperature and velocity range conditions may exist in which an exchanger could be run continuously without progressive deposition occurring and that the suspended solid could be removed at a later stage by filtration. Such a process is given added pertinence by the development of a successful filtering device (Hall and Haselden 1947) for ice crystal suspensions.

Any attempt to account theoretically for these results is made uncertain by the lack of information on the precise mechanism by which the production and growth of crystals takes place. Two mechanisms are possible: (i) All condensation.

takes place at the glass or metal interface and further growth of the crystals occurs chiefly whilst the crystals are in contact with the wall. When the crystals become too large they are partially or wholly eroded by the air stream. (ii) Condensation takes place by mixing of the cold boundary air layer with the turbulent main stream to produce small crystals suspended in the air stream. These crystals act as nuclei for further condensation which again occurs mostly at the interface between the boundary air flow and the main stream. Adhesion takes place, mainly due to thermal attraction of crystals to the cold surface.

For condensing liquids (Lewis 1927) substantial evidence is available to show that in the majority of cases process (i) is operative. In the given case, however, where the concentration of the condensable constituent in the gas phase is very small, and where the very open structure of the deposit has a large influence on the flow pattern of the air stream in the neighbourhood of the surface, it is likely that both processes are taking place. The relative importance of the two processes probably varies as the deposition proceeds, process (i) giving place to process (ii). It is also possible that migration of molecules over the surface of the deposited crystals can occur as shown by Estermann and Volmer (1927). Thus water vapour molecules may give up energy to the cooling surface and thereafter diffuse along the cooling surface and over the surface of the deposited crystals before taking a position in the crystal lattice. Evidence for mobility of molecules in ice crystals is given by Burton and Oliver (1935).

Further experiments, using ciné micrography, are now being undertaken to obtain more detailed information on the processes involved.

ACKNOWLEDGMENTS

The author wishes to acknowledge his indebtedness to Professor Sir Alfred Egerton for his interest in the work and to the Council of the Gas Research Board for sponsoring the work and giving permission to publish.

REFERENCES

- BURTON, E. F., and OLIVER, W. F., 1935, *Proc. Roy. Soc. A*, **153**, 166.
ESTERMANN, I., and VOLMER, M., 1921, *Z. Phys.*, **7**, 13.
HALL, T. A., and HASELDEN, G. G., 1947, *Proc. Inst. Chem. Eng.*, **25**, 132.
KOO, E. C., 1932, *Trans. Amer. Inst. Chem. Eng.*, **28**, 56.
KORNEMANN, H. C., and YENDALL, E. F., 1944, American Patent Specification 599,252.
LEWIS, W. K., 1927, *Trans. Amer. Inst. Chem. Eng.*, **20**, 9.
MCADAMS, W. H., 1942, *Heat Transmission* (New York and London : McGraw-Hill), p. 128.

A Relationship between the Refractive Index and the Infra-Red Threshold of Sensitivity for Photoconductors

By T. S. MOSS

Department of Colloid Science, Cambridge

Communicated by G. B. B. M. Sutherland; MS. received 26th May 1949 and in amended form 13th August 1949

ABSTRACT. It is shown that for photoconductive compounds the long wavelength 'threshold' of sensitivity λ should be related to the refractive index n of the photoconductor, and analysis of the available data shows that for the more highly refractive compounds n^4/λ always has a value close to 77. As a consequence of this relation, an explanation can be given for the change of the threshold wavelength with temperature which is observed for lead sulphide and similar materials. Suggestions are made as to what materials ought to be suitable for photoconductive detectors at longer wavelengths in the infra-red, and some confirmation of the ideas is shown by results obtained with cadmium arsenide, the properties of which are briefly described.

§ 1. INTRODUCTION

THERE are many materials now known for which the conductivity alters on exposure to radiation in the visible or near infra-red regions of the spectrum, but as yet there has been no explanation of the factors which determine the region of the spectrum to which a given material will be sensitive.

On the band theory of solids, insulators and semiconductors are materials for which the band of conduction levels is empty at the absolute zero of temperature and is separated from the filled band by a forbidden zone in which no levels exist for the pure material. Impurities or lattice imperfections, however, will give rise to levels in the forbidden zone.

The threshold of the photoconductive effect will occur at the wavelength where the quanta are just sufficient to raise electrons into the conduction band, or so near to the conduction band that thermal energies are sufficient to complete the process.

§ 2. THEORY

The general theory of photoconductivity worked out for F centres in alkali halides shows that absorption of the optical quantum first excites an electron without freeing it from the centre. The electron is then transferred to the conduction band by thermal energy from the lattice (Mott and Gurney 1940). It has often been suggested that even in so-called pure materials the photo-effect takes place only at preferred points in the lattice, such as vacant lattice sites, interstitial ions, or at atoms on the surface or in cracks, i.e. at any discontinuity in the lattice (Smekal 1929, Zwicky 1929, Gudden and Pohl 1926).

Any of these imperfections will create potential holes in the lattice capable of trapping electrons. An electron trapped at such a potential hole will be somewhat similar to an electron in an isolated atom, except that it is immersed in a medium of dielectric constant equal to that of the bulk material. Thus all the energy levels of the electron will be scaled down by a factor $1/(\kappa_{\text{eff}})^2$, where κ_{eff} is an 'effective' dielectric constant, which will approximate to the square of the refractive index (Mott and Gurney 1940). In principle the centre should have a line absorption spectrum, but in practice the lines are broadened into bands.

Hence the optical energy required to raise an electron at one of these irregularities into an excited state should be proportional to $1/\kappa_0^2$. We may thus expect that the threshold of the photoconductive effect, which is determined by the minimum energy required to raise an electron into an excited state, will vary inversely as the fourth power of the refractive index.

§ 3. EXPERIMENTAL DATA

Table 1 contains all the photoconductive compounds with refractive indices greater than about 2.5 for which both spectral sensitivity and refractive index data are available, with the exception of a few complex mixed crystals. According to Kurrelmeyer (1927) and Hughes and DuBridge (1932) the refractive index must be greater than 2 for photoconductivity to occur without artificially produced impurity centres.

§ 4. DISCUSSION

It will be seen that for the materials listed in Table 1 the values of n^4/λ are remarkably consistent. They all lie in the neighbourhood of 77 throughout a range of variation of n^4 from 35 to 230. For materials with refractive indices lying between 2 and 2.5 (mainly silver and thallium halides with barium, zinc and antimony oxides) the values of n^4/λ are somewhat lower and range from 72 for TlCl to between 47 and 58 for AgCl . Thus, although none of these materials gives results in complete disagreement with the relationship, it is clear that for lower values of the refractive index the variation of n^4/λ is considerable in view of the small range covered by n^4 . It may be noted that for n greater than 2.25 values of n^4/λ are greater than 65, indicating that the relation becomes progressively worse as n decreases.

The fact that the values are so similar for such a wide range of compounds indicates that the photoelectrons are originating from the same type of lattice imperfection in all cases. Thus the centres cannot be formed from impurities, with the possible exception of oxygen, which may be present in any materials, even if manufactured with the most careful vacuum technique.

The most probable type of centre is an F centre, i.e. a missing negative ion replaced by an electron; or possibly an F' centre if the potential hole has two bound electrons. According to Mott and Gurney (1940), the analogous centre consisting of a missing positive ion with a bound 'positive hole' (i.e. an electron missing from one of the adjacent negative ions) is also possible. These types of centre are just as likely to occur in any of the materials.

Pearson and Bardeen (1949) have recently carried out calculations on the excitation energies of boron impurities in silicon, assuming the energies to be sealed down by a factor κ^2 , and obtained values in reasonable agreement with the thermally measured activation energies.

For photoconductive elements, which are not ionic crystals and where the photoconductivity is intrinsic, the relationship does not apply, and the values of n^4/λ vary widely.

It is concluded, therefore, that for the more refractive photoconductive compounds there is some theoretical basis and experimental evidence for an approximate relation between the 'threshold' wavelength and refractive index, namely $n^4/\lambda = 77$.

Table 1

Material	n	λ	n^4/λ	References	
				Refractive index	Wavelength
PbO	2.53, 2.66	0.53	85	Landolt-Bornstein	SIGESO 1945
HgS	2.58, 2.82	0.63	84	{ Landolt-Bornstein Int. Critical Tables	Gudden and Pohl 1920
	2.76, 3.06		113		
HgO	2.37, 2.5, 2.65	0.55	71	Landolt-Bornstein	Dechene 1939
PbCrO ₄	2.5	0.55	71	Kapp 1935	Kapp 1935
Cu ₂ O	2.56	0.63	68	Landolt-Bornstein	Faltz 1937
Ag ₂ S	3.15 estimated	1.35	73	See note (e)	Coblentz and Kahler 1922
Sb ₂ S ₃	2.9 approx.	0.8	88	Billings and Hyman 1947	Coblentz 1920
MoS ₂	5.57, 2.33	2.0	84	{ Bailly 1938 Coblentz and Kahler 1919	Coblentz and Kahler 1919
	3.3		59		
CdS	2.43, 2.46	0.54	66	Landolt-Bornstein	Gudden and Pohl 1920
PbS	3.9	2.8	83	Handbk. Chem. Phys. 1948	Moss 1947
Tl ₂ S	2.9 estimated	1.0	71	See note (e)	Hughes and DuBridge 1932
Average			77		

Notes on the Table.

(a) It would be preferable to quote the refractive index in the region of the spectrum where the material is transparent and the dispersion small. However, few of the materials quoted have been measured over a large range of wavelengths, and so for consistency the index has been quoted for red light (0.65 to 0.75 μ) whenever possible. For most materials this wavelength should be in the non-absorbing region.

(b) For non-isotropic media with more than one refractive index the geometric mean has been used to calculate n^4/λ .

(c) Most of the measurements have been carried out on natural crystals, and both n and λ often vary from specimen to specimen, or as reported by different observers.

(d) In order to be applicable to the various types of spectral sensitivity curves which have been reported, the 'threshold' wavelength λ is taken as the point where the sensitivity finally begins to decrease rapidly. When the spectral curve is sharply peaked this wavelength is taken as the wavelength of maximum sensitivity. For materials where the sensitivity is roughly constant over a range of wavelengths, PbS for example, it is the point where the exponential drop begins on the long wavelength side.

(e) The refractive index of Ag₂S could not be found in the literature, but a fairly good estimate was obtained from data on reflectivity (Schneiderhohn 1931), using $R = (n-1)^2/(n+1)^2$, which gives a value of n equal to 3.4.

A second estimate was obtained from the refractive index of proustite, which is largely Ag₂S. It has the composition 3Ag₂S. As₂S₃ and $n = 2.99, 2.72$, or a mean value of 2.85 (Landolt-Bornstein Tables). Since for As₂S₃ $n = 2.6$ (Landolt-Bornstein Tables), the refractive index of Ag₂S should be 2.9 approximately. The mean of these two estimates, namely 3.15, has been used.

Similarly, for Tl₂S, reflectivity data give $n = 3.0$ (von Hippel *et al.* 1946). Also, for Tl₂S. As₂S₃, $n = 2.72$ (Dana 1944). With $n = 2.6$ for As₂S₃, as above, we obtain for Tl₂S $n = 2.8$. Hence the mean value is $n = 2.9$.

(f) Schonwald (1932) has reported that for Cu₂O there is an additional band at longer wavelengths. It is of relatively low sensitivity and has not been included in the data of Table 1.

§5. TEMPERATURE VARIATION OF THRESHOLD WAVELENGTH

Since the threshold wavelength is determined mainly by the refractive index of the material, it is reasonable to suggest that the shift of the threshold to longer wavelengths on cooling, as observed for PbS, PbTe etc., is due to a temperature-variation of the refractive index.

Unfortunately there seem to be no published data on the temperature variation of the refractive index for highly refractive materials. However, a rough estimate may be made in the following manner.

The Lorentz-Lorenz formula for the dielectric constant at optical frequencies is

$$\frac{\kappa_0 - 1}{\kappa_0 + 2} = CN\alpha,$$

where N is the number of molecules per unit volume, α the polarizability of the ions, and C a constant. Differentiating,

$$\frac{d\kappa_0}{dT} = \frac{1}{3}(\kappa_0 - 1)(\kappa_0 + 2) \frac{1}{N} \frac{dN}{dT}.$$

Now $\frac{1}{3N} \frac{dN}{dT} = \frac{-1}{3V} \frac{dV}{dT}$ = linear expansion coefficient for the material, = $-\beta$ say.

Therefore $d\kappa_0/dT = -(\kappa_0^2 + \kappa_0 - 3)\beta = -\kappa_0^2\beta$ with sufficient accuracy. A similar calculation has been applied to the static dielectric constant by Bretsher (1934).

This relation shows that there will be an increase in κ_0 , and hence in λ , with decreasing temperature. Up to the time of writing, the only materials measured which gave an increase of threshold wavelength on cooling are the PbS, PbSe, PbTe series, whereas for other materials on which data are available, i.e. Sb_2S_3 (Coblentz 1920) and ZnS (Moglich and Rompe 1942), there is a slight decrease of wavelength on cooling.

That any increase on cooling would be expected to be small for these latter materials follows from the fact that a large value of $d\kappa_0/dT$ necessitates both high refractive index and high expansion coefficient.

These conditions are fulfilled for PbS, for which $\kappa_0 = (3.9)^2$ and $\beta = 20 \times 10^{-6}$ (*International Critical Tables*). Hence $d\kappa_0/dT = -46 \times 10^{-4}$. For Sb_2S_3 , however, although the refractive index is high, 2.9, β is small, namely 1.5×10^{-6} . Therefore $d\kappa_0/dT = -1.1 \times 10^{-4}$. For ZnS the variation of the photoconductive threshold with temperature has not been measured. However, it would be expected to behave similarly to CdS, which shows a slight decrease of λ on cooling, especially as the long-wave edge of the absorption band decreases slightly in wavelength (from 0.335μ to 0.325μ) in a manner similar to that observed in CdS (Moglich and Rompe 1942).

For ZnS , $\kappa_0 = (2.36)^2$ and $\beta = 6.7 \times 10^{-6}$, hence $d\kappa_0/dT = 2.3 \times 10^{-4}$.

It is thus clear that while the effect of changing dielectric constant may be predominant in determining the temperature variation of λ for PbS, it will contribute little in the cases of Sb_2S_3 and ZnS.

The fact that there is a small change in the opposite direction for these latter materials is presumably due to some secondary process operating in the reverse direction, which becomes predominant when the dielectric change is sufficiently small. Such a process has been considered by Radowsky (1948), who has shown that there is a broadening of the electron energy levels due to interaction with the vibrating lattice. Thus increasing temperature broadens both the full and

conduction bands of levels, and hence reduces the width of the forbidden zone. It is shown that the effect gives approximately the correct shift of the absorption edge in ZnS.

The magnitude of the wavelength change expected may be evaluated for PbS. For the temperature interval 290° to 90° K.,

$$\Delta\kappa_0 = \kappa_0^2 \beta \Delta T = 0.92.$$

Therefore

$$\frac{\lambda_{90}}{\lambda_{290}} = \left\{ \frac{\kappa_0 + \Delta\kappa_0}{\kappa_0} \right\}^2 = 1.13,$$

corresponding to a change from 3.0 to 3.4μ . Although this is smaller than the observed values, it shows that the theory predicts for PbS a large increase of wavelength on cooling, in contrast to the small values of the wavelength changes when of the opposite sign. The limiting values observed by the author for PbS are: minimum 2.8μ to 3.4μ , maximum 2.9μ to 3.8μ .

The Lorentz-Lorenz formula is only an approximation, particularly for highly refractive materials, and a better check on the theory may be made by comparing the results for PbS and PbTe, so that any deficiencies in the above formula tend to be the same for both materials.

If E is the quantum energy corresponding to the threshold wavelength λ , then $E\kappa_0^2 = \text{constant}$, so that

$$\kappa_0^2 \frac{dE}{dT} = \frac{-2\kappa_0 E d\kappa_0}{dT} = 2E d\kappa_0^3 \beta.$$

Hence $(dE/dT)E^{-1}\beta^{-1} = \text{constant}$.

Values of this expression for PbS and PbTe are shown in Table 2. The agreement is seen to be satisfactory.

Table 2

	λ_{290}	Shift $\lambda_{90} - \lambda_{290}$	$\beta \times 10^{-6}$	$\frac{dE}{dT} E^{-1} \beta^{-1}$
PbS	2.8	0.6	20	30
	2.9	0.9	20	38
PbTe	3.5*	1.5	27	33

* Moss (1948) extrapolated to room temperature.

Note: All the threshold wavelengths are taken at the point where the sensitivity has fallen to 50%. The value of β for PbTe was measured by the author.

§ 6. DESIGN OF PHOTOCONDUCTORS FOR LONGER WAVELENGTHS

The relation found between wavelength of the photoconductive effect and the refractive index gives an indication of where to search for materials which will be photoconductive at longer wavelengths.

It is now possible to make PbS and PbTe detectors whose performance is more than a hundred times better than conventional infra-red detectors (such as thermopiles and bolometers), for comparable conditions of operation (Watts 1949, Simpson *et al.* 1948). This means that these photoconductors are superior to any other infra-red detectors within their spectral range, and it is thus of importance to extend this range to longer wavelengths. To this end the foregoing relation should be of value in shortening the tedious process of producing photoconductivity by trial and error in the many possible compounds so far uninvestigated.

It may be objected that there are no published data on materials of refractive index greater than that of PbS, and that it is not easy to measure such high refractive indices. However, if the above ideas are correct, considerable guidance as to where to search in the periodic table for new materials can be obtained by considering data on the polarizability of the ions. The higher the polarizability (with other things equal) the greater will be the refractive index, and consequently the wavelength for which the material may be photoconductive. The data on some of the comparable groups of photoconductors are given in Table 3. They are arranged in general so that the 'threshold' wavelength increases diagonally towards the lower right-hand corner. The polarizability of each ion is quoted beside the name of the ion, and the wavelengths (in microns) for various combinations of the ions given. Pauling's (1927 b) figures are used for the polarizability of the ions, except for Tl^+ from Wulff and Schaller (1934), and Pb^{++} from Wulff and King (1934).

Table 3

		Cl^-	Br^-	I^-	O^{--}	S^{--}	Se^{--}	Te^{--}
	($\times 10^{-24}$)	3.6	4.8	7.1	3.9	10	10.5	14
Zn^{++}	0.3				$0.31\mu^{(1)}$	$0.4\mu^{(2)}$		
Cd^{++}	1.1					0.54μ	$0.7\mu^{(3)}$	$0.8\mu^{(3)}$
Hg^{++}	1.26			$0.55\mu^{(4)}$	0.55μ	0.63μ		
Sb^{+++}					$0.32\mu^{(8)}$	0.8μ		
Bi^{+++}						$1.0\mu^{(5)}$	$1.1\mu^{(6)}$	1.6μ or $2.8\mu^{(7)}$
Tl^+	3.7	$0.36\mu^{(9)}$	$0.41\mu^{(9)}$	$0.45\mu^{(9)}$		1.0μ		
Ag^+	1.7	$0.38\mu^{(10)}$	$0.43\mu^{(10)}$	$0.42\mu^{(10)}$		1.35μ		
Pb^{++}	3.7			$0.52\mu^{(9)}$	0.53μ	2.8μ	$3.6\mu^{(11)}$	$3.6\mu^{(11)}$

References in Table: (1) Veselovski 1948; (2) Gudden and Pohl 1920; (3) Frerichs 1947; (4) Volmer 1917; (5) Coblenz and Kahler 1919, Asai 1940; (6) SIGESO 1945; (7) Measured by author; (8) Heintz 1923; (9) Coblenz and Eckford 1923; (10) Kirillov 1929; (11) Moss (1949 b), room temperature values.

The polarizabilities of the Sb^{+++} and Bi^{+++} ions are not known, but qualitatively the polarizability increases with atomic weight, so that Bi^{+++} will be the greater.

The wavelengths for the oxides and iodides are very similar, although the I^- ion has considerably greater polarizability. This is due to the fact that the I^- ion is large (radius 2.2 Å. compared with only 1.3 Å. for the O^{--} ion), with consequent reduction in the polarizability per unit volume. The I^- ion is almost identical in size with the Te^{--} ion, and hence a direct comparison can be made for these two.

It will be seen that the longest wavelength 'threshold' so far obtained is that for PbTe .

The superiority of the Pb^{++} ion may be appreciated from the fact that it has the largest polarizability quoted in the literature, and has also a relatively small radius (1.2 Å.) compared with Tl^+ (1.44 Å.) for example. Furthermore, the ion is divalent, and hence in the case of the telluride a larger fraction of the lattice space is filled by the more highly polarizable Te^{--} ion, compared with Tl_2Te , which increases the polarizability per unit volume.

Among the ions listed (or similar ones in the periodic table) the Te^{--} ion is clearly the most suitable negative ion, and it is doubtful if a positive ion superior to the Pb^{++} ion can be found. The most hopeful materials to try would seem to be Ag_2Te , Tl_2Te and Au_2Te , though it is almost certain that all three would be somewhat inferior to PbTe . The Au^+ ion has a polarizability 1.9×10^{-24} and radius 1.37 Å. (Goldschmidt 1927, Pauling 1927a, for ionic radii).

The metals in the first four columns of the periodic table with their normal valencies (i.e. Li^+ , Be^{++} , B^{+++} , C^{++++} and their congeners) do not seem very promising. No figures are available for the molybdenum-tungsten series, which might be worthy of investigation. Further work should also be done on bismuth compounds, since Bi_2S_3 has been described as possessing some sensitivity out to 7.0μ (United States Patent Specification 1941). What seems a much more fruitful field for research is the investigation of the completely untried series of compounds with trivalent negative ions. Table 3 shows that for ions of the same size (e.g. I^- and Te^{--}) the divalent ion has much higher polarizability and results in much longer wavelengths. For comparison, the As^{---} ion, which is the same size as the I^- and Te^{--} ions, has the very high polarizability of 28×10^{-24} . If, therefore, photoconductive layers of arsenides could be prepared, a definite increase in the wavelength range would be expected.

Photosensitive layers of one such material have now been prepared, viz. cadmium arsenide, Cd_3As_2 . The spectral sensitivity curve has been measured, and the 'threshold' wavelength is found to be beyond 2.0 microns. This is a very large increase over the wavelength of 0.8μ for CdTe , particularly compared with the small increase resulting on going from CdS to CdTe . The properties of this material are described in the Appendix.

This result is adequate confirmation of the idea that the arsenides as a class should be sensitive to longer wavelengths. If sensitive layers of lead arsenide (and possibly silver or thallium) can be made, there is every hope that they will extend the present range of photoconductive detectors further into the infra-red region.

ACKNOWLEDGMENTS

Acknowledgment is made to the Chief Scientist, Ministry of Supply, and to the Controller of H.M. Stationery Office for permission to publish this paper.

The author wishes to thank Dr. G. B. B. M. Sutherland for his advice and encouragement in the work.

REFERENCES

- ASAI, T., 1940, *Bull. Inst. Phys. Chem. Res.*, **19**, 1403; 1941, *Ibid.*, **20**, 1.
 BAILLY, R., 1938, *Bull. Acad. Belg. Cl. Sci.*, **24**, 791.
 BILLINGS, B. H., and HYMAN, M., 1947, *J. Opt. Soc. Amer.*, **37**, 119.
 BRETHER, E., 1934, *Trans. Faraday Soc.*, **30**, 684.
 COBLENTZ, W. W., 1920, *Bur. Stand. Bull., Wash.*, **16**, 595; 1923, *Ibid.*, **18**, 585.
 COBLENTZ, W. W., and ECKFORD, J. F., 1923, *Bur. Stand. Bull., Wash.*, **18**, 489.
 COBLENTZ, W. W., and KÄHLER, H., 1919, *Bur. Stand. Bull., Wash.*, **15**, 121, 231; 1922, *Ibid.*, **18**, 265.
 DANA, —, 1944, *Dana's System of Mineralogy*, 7th Edition (New York: Wiley).
 DECHENE, G., 1939, *C.R. Heb. Ac. Sci. Inst. France*, **208**, 95.
 FALTZ, G., 1937, *Ann. Phys., Lpz.*, **30**, 193.
 FRERICHS, R., 1947, *Phys. Rev.*, **72**, 594.
 GOLDSCHMIDT, V. M., 1927, *Chem. Ber.*, **60**, 1263.

- GUDDEN, B., and POHL, R. W., 1920, *Z. Phys.*, **2**, 181, 361; 1926, *Ibid.*, **37**, 881.
Handbook of Chemistry and Physics, 1948 (Ohio : Chemical Rubber Publ. Co.).
 HEINTZ, W., 1923, *Z. Phys.*, **15**, 339.
 HINTERBERGER, H., 1942, *Z. Phys.*, **119**, 1.
 VON HIPPEL, A., *et al.*, 1946, *J. Chem. Phys.*, **14**, 355.
 HUGHES, A. L., and DUBRIDGE, L. A., 1932, *Photoelectric Phenomena* (New York : McGraw-Hill Book Co.).
International Critical Tables, 1926 (New York : McGraw-Hill Book Co.).
 KAPP, G., 1935, *Ann. Phys., Lpz.*, **22**, 257.
 KIRILLOW, E. A., 1929, *Z. Wiss. Photogr.*, **26**, 235.
 KURRELMMEYER, B., 1927, *Phys. Rev.*, **30**, 901.
 LANDOLT-BORNSTEIN, or LANDOLT-BORNSTEIN-ROTH-SCHEELE, 1912, 1923, 1927, 1931, *Physical Chemical Tables* (Berlin : Springer).
 MOGLICH, F., and ROMPE, R., 1942, *Z. Phys.*, **119**, 472.
 MOSS, T. S., 1947, *Nature, Lond.*, **159**, 476; 1948, *Ibid.*, **161**, 776; 1949 a, *Proc. Phys. Soc. A*, **62**, 264; 1949 b, *Proc. Phys. Soc. B*, **62**, 741.
 MOTT, N. F., and GURNEY, R. W., 1940, *Electronic Processes in Ionic Crystals* (Oxford : Clarendon Press).
 PAULING, L., 1927 a, *J. Amer. Chem. Soc.*, **49**, 765; 1927 b, *Proc. Roy. Soc. A*, **114**, 181.
 PEARSON, G. L., and BARDEEN, J., 1949, *Phys. Rev.*, **75**, 865.
 RADOWSKY, A., 1948, *Phys. Rev.*, **73**, 749.
 SCHNEIDERHOHN, H., 1931, *Lehrbuch der Erzmikroskopie*, Vol. 2 (Berlin : Borntraeger).
 SCHONWALD, B., 1932, *Ann. Phys., Lpz.*, **15**, 395.
 SIGESO, 1945, *Government report on German wartime work*. SIGESO 56, Section 1, Paras. 172 and 173.
 SIMPSON, O., SUTHERLAND, G. B. B. M., and BLACKWELL, D. E., 1948, *Nature, Lond.*, **161**, 281.
 SMEKAL, A., 1929, *Z. Phys.*, **55**, 289.
 UNITED STATES PATENT SPECIFICATION, 1941, U.S. Patent No. 2,406,139.
 VESELOVSKI, V. I., 1948, *Zhur. Fiz. Khim.*, **22**, 1427.
 VOLMER, M., 1917, *Z. Wiss. Photogr.*, **16**, 152.
 WATTS, B. N., 1949, *Proc. Phys. Soc. A*, **62**, 457.
 WULFF, P., and KING, A., 1934, *Z. Kristallogr.*, **87**, 72.
 WULFF, P., and SCHALLER, D., 1934, *Z. Kristallogr.*, **87**, 43.
 ZWICKY, F., 1929, *Proc. Nat. Acad. Sci., Wash.*, **15**, 816.

APPENDIX

PROPERTIES OF CADMIUM ARSENIDE

The arsenides represent a class of semiconductors and photoconductors which have not as yet been investigated.

(i) *Preparation.*

The compound Cd_3As_2 was prepared by direct fusion of the elements. The materials were weighed out with a slight excess of arsenic and inserted into a Pyrex tube. The tube and contents were outgassed *in vacuo* at 160°C. and then sealed off from the system (pressure 10^{-6} mm. Hg approx.). The tube was baked at 610°C. for two hours. A few milligrams of the resulting material were used in the manufacture of a typical layer, the method of preparation being similar to that previously used for tellurium layers (Moss 1949 a).

(ii) *Resistance.*

The resistance of one layer was measured over the temperature range 90° to 390° K. The results are plotted as log (resistance) against reciprocal temperature in Figure 1. The curve is roughly linear between 390° and 180° K., and corresponds

to $R = R_0 \exp(\epsilon/2kT)$, where $\epsilon = 0.14$ ev. Presumably the slope of the line would be steeper if the material could be made more accurately stoichiometric, in a similar manner to PbS (Hinterberger 1942). The resistance was substantially constant below 140°K .

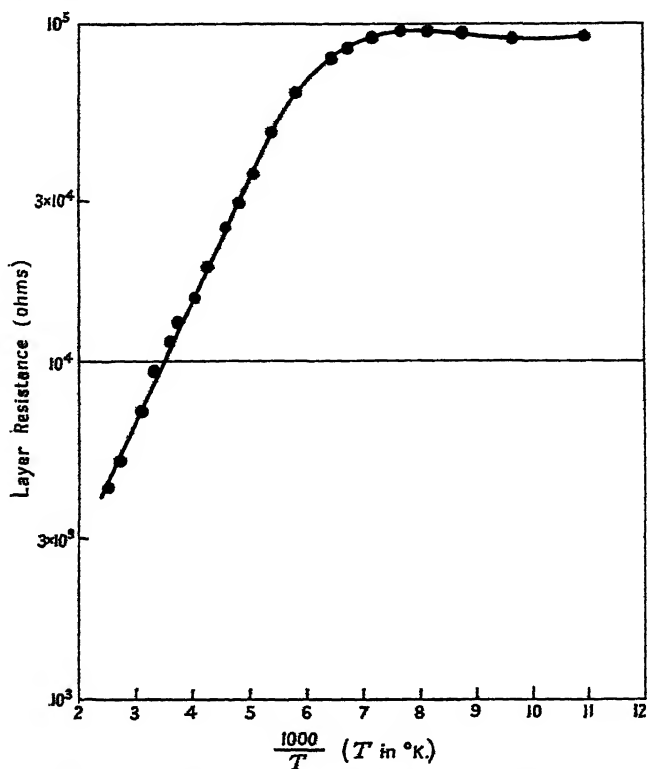


Figure 1. Resistance temperature relation for Cd_3As_2 .

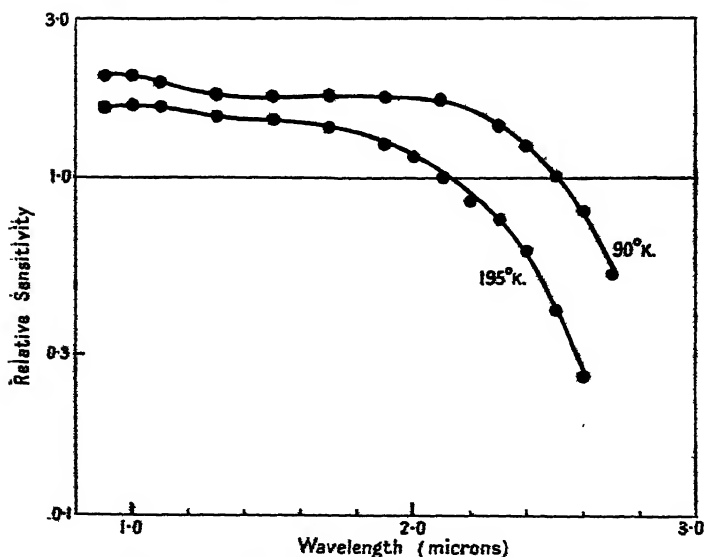


Figure 2. Spectral sensitivity curves for Cd_3As_2 .

At room temperature the resistance was 10^4 ohms (the layer dimensions being 1 mm. gap between 3 mm. electrodes and approximately 10^{-3} mm. thick), which corresponds to a specific resistance of 3 ohm cm.

(iii) *Type of current carriers.*

For all samples thermo-electric power measurements indicated that the carriers were electrons.

(iv) *Photoconductivity.*

The layers were sensitive only when cooled. The response times, measured as the time required for the electrical signal to decay to $1/e$ on removal of the light signal, lay in the range 10^{-3} to 4×10^{-3} second.

The spectral distribution of the photosensitivity was measured, using a monochromator with a lithium fluoride prism. The a.c. signal from the layer produced by the chopped radiation from the monochromator was passed through a high gain amplifier tuned to the chopping frequency and read on a rectifier type meter. The signal output, referred to a constant energy spectrum, is plotted against wavelength in Figure 2 for two layer temperatures, 90° and 195° K.

It will be noted that the wavelength at which the sensitivity starts to fall off increases as the temperature is lowered, in a manner similar to PbS. The wavelengths at which sensitivity has fallen to 50% of its value on the flat portions of the curves (i.e. 1.5μ region) are: 90° K., 2.54μ ; 195° K., 2.3μ . The value of the coefficient of linear expansion has been measured and found to be $\beta = 17 \times 10^{-6}$ per $^\circ$ C.

Hence $(dE/dT)E^{-1}\beta^{-1} = 36$, in good agreement with the values for PbS and PbTe shown in Table 2.

Capacitance Measurements on Selenium Rectifiers: Evidence of Anomalous Dispersion

By R. COOPER

Associated Electrical Industries, Ltd., Aldermaston, Berks.

MS. received 22nd July 1949

ABSTRACT. The paper presents evidence showing that, contrary to the generally accepted view, the capacity of a selenium rectifier at audio frequencies may vary with the frequency of measurement. The magnitude of the variation depends on the metal used as counter-electrode. Thus in the frequency range 300–2,000 c/s. little variation occurs with antimony or bismuth, but with cadmium there is a considerable decrease in capacitance. The effect is considered due to polarization arising from the relative motion in the selenium lattice of the negatively charged impurity centres and positive ions which diffuse into the selenium from the counter-electrode.

It has previously been established (Schottky 1942) that a selenium rectifier may be represented by the circuit illustrated by Figure 1. In this circuit C represents the equivalent capacitance of the barrier layer, R its differential resistance and r the resistance of the selenium layer between the two electrodes. The method generally used to determine the components of this circuit is to find

the effective series resistance and reactance of the rectifier at various frequencies, the measurements being made with a small alternating voltage (less than 20 mv.) superimposed on a constant D.C. bias voltage. When the series reactance is plotted against the series resistance, a semicircular impedance locus diagram is obtained with its centre on the resistance axis. The diameter of the semicircle is equal to the barrier resistance R , and its smallest intercept with the resistance axis yields the selenium layer resistance r . The barrier capacitance C at any frequency f is given by

$$C = \frac{B}{A-r} \frac{1}{R\omega},$$

where B is the rectifier's equivalent series reactance, A its equivalent series resistance and ω the angular frequency ($\omega = 2\pi f$). Both R and C depend on the D.C. bias voltage applied to the rectifier.

It follows from the semicircular locus diagram that R and r must be independent of frequency at least within limits of experimental error. It is generally supposed that C is also independent of frequency, but this does not follow directly from the existence of a semicircular locus diagram: variation of C with frequency will not cause deviation of points from the semicircle but only their shift along its circumference. The circle diameter depends only on the barrier resistance R .

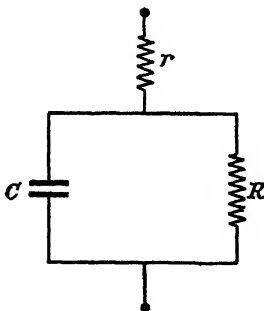


Figure 1. Equivalent circuit of selenium rectifier.

At a particular frequency the value of C determines the angle θ in Figure 3 ($\tan \theta = B/A$) and hence the point where the tip of the impedance vector Z touches the circumference of the semicircle.

Experiments made in this Laboratory have shown the circuit of Figure 1 to represent the behaviour of selenium rectifiers with counter-electrodes of bismuth and antimony. In both cases, analysis of the measurements in the frequency range 300–2,000 c/s. shows the barrier capacitance C to be sensibly independent of frequency. This is not the case, however, with rectifiers possessing cadmium counter-electrodes. The impedance locus is generally not a semicircle with its centre on the resistance axis, and thus the circuit of Figure 1 cannot be applied. However, if the D.C. bias voltage is adjusted to cause the D.C. resistance R of the barrier to decrease, the usual form of rectifier semicircular locus diagram is eventually realized. This occurs when the barrier resistance is less than about 2,000 ohms. Analysis of such a semicircular locus diagram indicates the barrier capacitance C to be a function of frequency.

Figure 2 illustrates the variation with frequency of the capacitance of a rectifier possessing a cadmium counter-electrode (area 1.3 cm^2). The three curves correspond to

- (a) the rectifier at 50°C. with $+0.1 \text{ v. D.C. bias.}$
- (b) the same rectifier at 20°C. with $+0.1 \text{ v. D.C. bias.}$
- (c) the rectifier at 50°C. with $+0.04 \text{ v. D.C. bias.}$
- (d) the rectifier at 20°C. with $+0.04 \text{ v. D.C. bias.}$

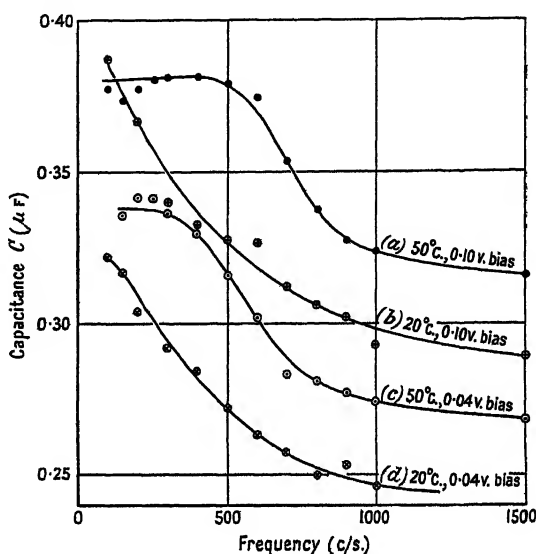


Figure 2. Variation of capacity with frequency for a selenium rectifier with a cadmium counter-electrode.

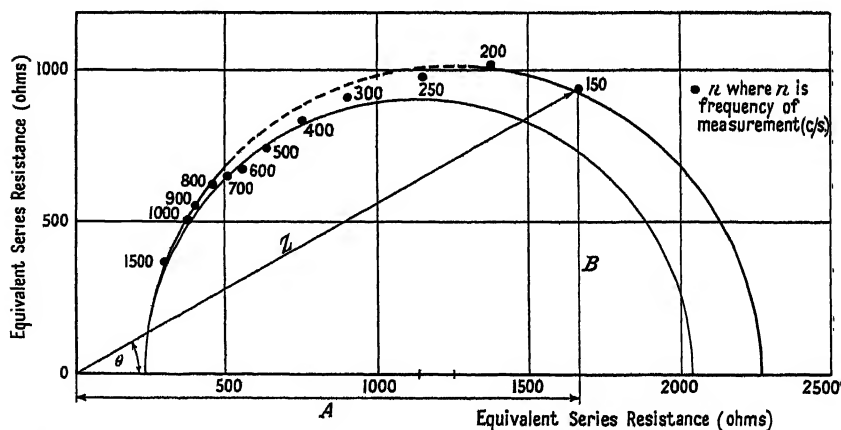


Figure 3. Impedance locus diagram. Bias volts = $+0.04$ volt, temperature = 50°C.

Figure 3 illustrates the locus diagram corresponding to (c). The impedance points in the range 250–700 c/s. do not lie precisely on the same circle as those at frequencies up to 250 c/s. and above 800 c/s. There appears to be a decrease in the value of R in the region 300–700 c/s. since the corresponding points lie on a circle (or circles) of smaller diameter. This can be accounted for by assuming the existence, in parallel with R , of a resistance R_f which is a function of frequency.

The value of R_f is comparable with R in the frequency range 250–700 c/s. but at other frequencies it is appreciably greater than R . Reference to Figure 2 shows that the above frequency range is the one in which the rectifier capacitance rapidly decreases. Therefore in order to make the circuit of Figure 1 correctly to represent the rectifier, a second resistance R_f must be added in parallel with R and C . The behaviour of C and R_f with respect to increase of frequency will be recognized as characteristic of a dielectric exhibiting anomalous dispersion, and consistent with this is the effect, exhibited in Figure 2, of temperature increase on the frequency band in which the change of capacitance occurs.

Anomalous dispersion in a dielectric is usually described by the decrease of capacitance or dielectric constant of the condenser with increasing frequency, and the accompanying changes in the tangent of the loss angle δ . For a lossy dielectric represented by a capacity C in parallel with a resistance R_f , $\tan \delta = 1/R_f C \omega$. The information contained in Figures 2 and 3 enables an estimate to be made of the maximum loss angle of the rectifier due to anomalous dispersion. The measurements made do not exactly indicate the frequency corresponding to the maximum loss angle, but in the case of the rectifier at 50°C. with 0.04 volt D.C. bias, the circle diagram of Figure 3 indicates 600 c/s. to be a reasonable value. At this frequency the capacitance was $0.3 \mu F$. The corresponding value of R_f can be estimated from the difference in diameter of the two semicircles shown in Figure 3 and was approximately 2.0×10^4 ohms. These figures yield a value of 0.044 for $\tan \delta_{\max}$ at 600 c/s.

At present the mechanism of the dispersion process is not fully understood but the fact that it occurred in the frequency range 200–2,000 c/s. with cadmium, but not with antimony or bismuth, shows that it is not a property of the spectroscopically pure selenium used in the preparation of the rectifiers. Recent work by Breckenridge (1948), who has observed anomalous dispersion at audio frequencies in alkali halide crystals, suggests a possible mechanism. This is that polarization results from relative motion in the selenium lattice of the negatively charged impurity centres (responsible for the defect electron conduction in selenium) and the positively charged metal ions which diffuse into the selenium from the counter-electrode.

Following Breckenridge, and applying the calculation given by Debye for anomalous dispersion and absorption in a solid, a second estimate of the maximum loss angle δ_{\max} can be made. Debye (1945) showed that

$$\tan \delta_{\max} = (\epsilon_1 - \epsilon_0) / 2 \sqrt{(\epsilon_1 \epsilon_0)},$$

where ϵ_1 is the static or low-frequency dielectric constant, and ϵ_0 is the dielectric constant at frequencies very much higher than those at which anomalous dispersion occurs. Using values from Figure 2 curve (c), the above equation yields 0.115 for $\tan \delta_{\max}$. This is in reasonable agreement with the value of 0.044 previously estimated.

REFERENCES

- BRECKENRIDGE, R. G., 1948, *J. Chem. Phys.*, **16**, 959.
 DEBYE, P., 1945, *Polar Molecules* (New York: Dover Publications), p. 102.
 SCHOTTKY, W., 1942, *Z. Phys.*, **118**, 539.

On Dissociation Processes in Certain Gases of High Dielectric Strength

BY J. W. WARREN*, W. HOPWOOD† AND J. D. CRAGGS†

University of Liverpool

MS. received 31st August 1949

ABSTRACT. The dissociation processes, due to electron collisions, in certain polyatomic gases such as CCl_4 , CCl_2F_2 are discussed with reference to the unusually high dielectric strength of certain of these gases. Possible explanations of the latter are then suggested in terms of these dissociation processes.

§ 1. INTRODUCTION

THE high dielectric strength of certain gases, such as CCl_4 , CCl_2F_2 , SF_6 , relative to that of, say, nitrogen, was first noticed in 1889 by Natterer (see Charlton and Cooper 1937), and has been a subject of considerable interest for some years (Joliot, Feldenkrais and Lazard 1936, Rodine and Herb 1937, Charlton and Cooper 1937, Pollock and Cooper 1939, Hochberg and Sandberg 1942). It is also a matter of some technical importance. As an example, it was found (Hochberg and Sandberg 1942) that the dielectric strength of CCl_4 vapour in certain circumstances was about 6.4 times that for air. Some relevant data are collected in the following Table.

Relative Dielectric Strengths of certain Molecular Gases

Gas	(1)	(2)	(3)
CCl_4	—	6.36	6.3
CCl_2F_2	3.0 (725 mm. Hg)	4.47	3.4-4.4
CCl_2F_2	2.4	2.56	2.4-2.5
CF_4	—	—	1.1
Cl_2	1.55	—	—
SF_6	—	2.49	2.3-2.5

- (1) Dielectric strength relative to $\text{N}_2 \equiv 1$; uniform field gap, 1 atm. (Charlton and Cooper 1937).
- (2) Dielectric strength relative to air $\equiv 1$; uniform field gap by extrapolation, using Paschen's law, to 1 atm. (Hochberg and Sandberg 1942).
- (3) Dielectric strength relative to $\text{N}_2 \equiv 1$; as (2) (Hochberg and Sandberg 1946).

Various authors (Charlton and Cooper 1937) have endeavoured to connect the dielectric strength of such gases, as exemplified by the breakdown voltage of standard spark gaps, with various physical parameters such as molecular weight, but with little success, and no satisfactory explanation of the behaviour of the above mentioned and other similar polyatomic molecular gases has as yet been given. It has been widely known for some years that the presence of halogen atoms in the molecules of a gas is connected with unusually high dielectric strengths, as shown, for example, in the tables of data compiled by Charlton and Cooper (1937)

* Department of Physics.

† Department of Electrical Engineering.

and Hochberg and Sandberg (1942). Typical experimental work on electrical breakdown (corona or sparks) in such gases has been carried out by Charlton and Cooper (1937) and workers there cited, Pollock and Cooper (1939) (CCl_2F_2 , SF_6), Hochberg and Sandberg (1942, 1946) (many gases), Craggs and Meek (1948) (CCl_2F_2), and Weissler and Mohr (1947) ($\text{CCl}_2\text{F}_2/\text{air}$), but possible reasons for the interesting behaviour of these gases are usually only vaguely given. Some authors (Weissler and Mohr 1947, Mohr and Weissler 1947) suggest that the free halogens (F and Cl in the case of CCl_2F_2) which are undoubtedly produced by dissociation in electrical discharges are important in that the high electron affinity of these elements (Massey 1938) will lead to the formation of stable negative ions and a resulting decrease in the free electron population in the discharge.

It is believed that these suggestions can be supplemented or perhaps replaced after a consideration of other measurements on the halogen-containing molecular gases, but it is first necessary to discuss briefly two other general lines of investigation of their behaviour.

§2. PREVIOUS WORK ON GASES OF HIGH DIELECTRIC STRENGTH

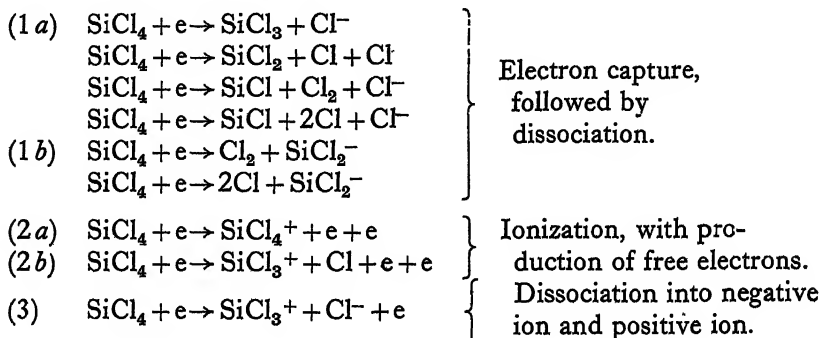
Firstly, Hochberg and Sandberg (1942, 1946) combined measurements of sparking voltages in SF_6 with measurements of Townsend's α coefficient. This is the principal controlling factor in the breakdown voltage (Loeb and Meek 1940), especially for breakdown in the streamer régime, i.e. for $pd > 200$ say, where p is the pressure in mm. Hg and d the discharge gap length in air. Thus, Meek's equation giving the breakdown field X_s in a spark gap for conditions in which the streamer mechanism applies is

$$X_s(d/p)^{1/2} = K\alpha e^{\alpha d}$$

for a uniform field gap. Hochberg and Sandberg show that the value of α in SF_6 is very low compared with that in air (at X/p 115–122, $\alpha(\text{SF}_6) \simeq 1/30 \alpha(\text{air})$) and proceed to suggest that the low value of α agrees reasonably with the high breakdown voltage in SF_6 (2.5 times that of air, in the conditions cited (Hochberg and Sandberg 1942)) on the basis of Meek's equation (Loeb and Meek 1940, Meek 1940). Hochberg and Sandberg (1946) later extended the measurements of α to include $\text{C}_2\text{H}_5\text{Cl}$, $\text{C}_2\text{H}_5\text{Br}$, C_5H_{12} , SF_6 (as a check on the earlier results), CHCl_3 and CCl_4 . The highest values of X/p for a given α/p occur for CHCl_3 and CCl_4 , and it is shown how high values of breakdown voltage are related to low values of α/p for all the gases studied. It is suggested that these low values of α are due to higher electron energy losses in inelastic non-ionizing collisions which should be more frequent with complex polyatomic molecules, but it then seems difficult to understand why relatively complex molecular gases, e.g. C_5H_{12} , should not have lower values of α than is the case.

Secondly, studies of dissociation processes due to electron collisions in several gases such as SF_6 (Dibeler and Mohler 1948), SiCl_4 (Vought 1947), SiF_4 (Dibeler and Mohler 1948), CCl_4 (Baker and Tate 1938, Warren and Craggs 1949), CF_4 (Dibeler and Mohler 1948, Warren and Craggs 1949), CClF_3 and CFCl_3 (Warren and Craggs 1949) have given results of considerable interest and complexity which have not previously been invoked in suggested explanations of the low value of the Townsend α in certain of these gases. Vought's study of the dissociation products of SiCl_4 appears to be the most complete analysis of its kind; he found the following reactions, amongst others, which may be of importance

in the present connection, and which are given here as examples of certain general types of collision processes (Massey 1938).



All these processes, especially (2 a), have long been known and studied, and Massey has given a particularly clear discussion of (1) and (3) with special reference to the case of oxygen. (1 a) can occur at very low energies (~ 1 ev.), particularly if the electron affinity of the atom forming the negative ion is high and if the energy required to dissociate the molecules is low. The process (3) usually requires an electron energy close to that for (2 a). There are, of course, other dissociation products formed at higher energies. It has been found that for all the gases listed in the last paragraph, except SiCl_4 , the reaction (3) is very much more probable than (2). Thus, for example, Dibeler and Mohler (1948) estimated that if I is the current, $I(\text{SF}_5^+)/I(\text{SF}_6^+)$ at 50–100 ev. was 5,000:1 and that $I(\text{SiF}_3^+)/I(\text{SiF}_4^+)$ was about 50:1 at 50–100 ev. Warren (Warren and Craggs 1950) was unable to detect $I(\text{CCl}_4^+)$ in a case where $I(\text{CCl}_3^+)$ was easily measurable. Baker and Tate (1938) quote similar data for CCl_4 and CCl_2F_2 . For the exceptional case of SiCl_4 (Vought 1947), $I(\text{SiCl}_3^+)/I(\text{SiCl}_4^+)$ was 100:56 at 75 ev. It has not been found possible to calculate the relative probabilities for reactions (2) and (3) in the present cases of polyatomic molecules.

§ 3. ELECTRICAL CHARACTERISTICS OF FLUOR- AND CHLOR-CARBONS

The low values of Townsend's α for SF_6 , CCl_4 etc. may be due partly to the molecular complexity and, therefore, to the greater possibility of inelastic non-ionizing collisions, although this cannot be the sole explanation since other complex gases give larger values of α . Free halogens are produced by dissociation (see reaction type (1)), and thus attachment can result with a resulting loss of free electrons, but this should also occur in Cl_2 and O_2 which show a 'normal' value of dielectric strength (Charlton and Cooper 1937, Stephenson 1933). Negative molecular ions can, of course, also occur in these simple gases. It would appear also that process (1), which occurs in O_2 (Massey 1938), is not primarily responsible for low values of α , although it must be made clear that the probability of electron capture (process (1)) might be unexpectedly great for the complex molecules. However, the relative scarcity of the primary ion (process (2 a)) with respect to the yield of higher products of ionization (process (3)), which has been observed for SF_6 , CCl_4 , etc., is a characteristic of this type of molecule (see, however, the work of Stevenson and Hipple 1942). It is to be expected that this might in turn lead to low values of α since, unless the electric field in the discharge is high enough to cause electron detachment by collision (Massey 1938), no contribution to a

Townsend avalanche can be made by a reaction of type (3). Thus the inefficient production of fresh electrons by process (2*a*) should lead to the observed low values of α , which in turn (Hochberg and Sandberg 1946) lead to high dielectric strengths. In the case of chlorine and fluorine negative ions the fields necessary to cause detachment would presumably need to be much greater than that corresponding to $X/p = 90 \text{ v/cm/mm. Hg}$ (Massey 1938), since this appears to be the necessary value in molecular oxygen which has an electron affinity probably about 0.2 e.v. and therefore appreciably less than that of chlorine or fluorine.

It will be noticed that there are large variations of dielectric strength and therefore of Townsend's α even amongst gases which are alike in that they show little tendency to form the primary ion; thus CF_4 (Hochberg 1947) has a dielectric strength close to that of N_2 , although (Dibeler and Mohler 1948) CF_4^+ is hardly detectable compared with CF_3^+ in the products of dissociation due to electron bombardment. Further, CCl_2F_2 and CCl_4 (Charlton and Cooper 1937, Weber 1942) have respectively about 2.5 and 6 times the dielectric strength of air or N_2 (for the particular conditions of the experiments cited), and both (Baker and Tate 1938) show very little tendency to form the primary ion (i.e. CCl_2F_2^+ or CCl_4^+). It is suggested that these variations may be due to the relative probabilities of reactions of types (2*b*) and (3), and that for the gases with the highest dielectric strength (CCl_4) reaction (2*b*), which would of course contribute to a Townsend α , is relatively improbable. Evidence is provided in support of this by the work of Baker and Tate (1938), whose results show that, if we make the apparently reasonable assumption (Massey 1938) that resonance capture of electrons, leading to reactions (1*a*) and (1*b*), is relatively negligible, the probabilities of the occurrence of reaction (3) compared with that of reaction (2*b*) in CCl_4 and CCl_2F_2 respectively are (at 75 e.v. electron energy) about 100:1 and 100:16. It seems possible, therefore, that reaction (2*b*) would still be more probable in CF_4 .

The reaction (2*b*) should be produced at a higher appearance potential than reaction (3), as Vought (1947) pointed out, by an amount, 3.8 e.v. equal to the electron affinity of, in this case, chlorine. Its onset was not noticed by Vought or mentioned by Baker and Tate, but this may be due to the large positive ion current found at about 3.8 e.v. above the appearance potential for (3). It is clear that measurements of negative and positive ion currents at energies just above the critical values are needed, and we propose to undertake this work.

The well-known fact (Charlton and Cooper 1937) that addition of CCl_4 or CCl_2F_2 (Weber 1942) to, say, N_2 gives an appreciably increased dielectric strength, is presumably due to the above mechanisms, since the critical potentials for these processes in CCl_4 or CCl_2F_2 are lower than the ionization potential for N_2 , and thus the electron temperature in any discharge in, say, N_2/CCl_4 is lowered compared with that in N_2 . It would thus appear possible that additions of SF_6 to N_2 , or of CCl_4 to Xe might not give rise to appreciably increased dielectric strength, and these experiments will shortly be carried out in these laboratories.

In conclusion, it may be stated that whilst the suggested mechanisms seem to explain qualitatively the known experimental data, it is essential to check the existing scanty information, particularly on negative ion yields at various bombarding electron energies, especially for reaction (3) as compared with reaction (2*b*), and to extend it. This work will be carried out in these laboratories and preliminary experiments on the appearance potentials for various reactions (Warren and Craggs 1949) have now been in progress since October 1948 in our

mass spectrometer laboratory. Further experiments on the dielectric strength of certain mixtures, on measurements of Townsend's α and on the investigation of dissociation products formed in electrical discharges are in hand in the Electrical Engineering Laboratories.

§ 4. SUMMARY

It is suggested that the dielectric strengths of the well-known chlor- and fluor-carbons, and of certain related compounds, such as SF_6 can be explained in terms of certain general dissociation reactions:



where X is usually an atom of carbon or sulphur and Y is a halogen atom, usually in the most useful gases, of chlorine or fluorine or combinations thereof. Examples of the latter are CClF_3 , CCl_2F_2 and CCl_3F , which we have studied, and in which the most important of reactions (B) and (C) are those in which the escaping Y atom is chlorine. It is known that the values of Townsend's α coefficient, which, in given geometrical conditions and particularly at high pressures, largely determines the dielectric strength of a discharge gap (Hochberg and Sandberg 1946), are very low for CCl_4 , CCl_2F_2 , SF_6 etc., and also that in such gases the reaction (B) is much more probable than reaction (A). It is suggested here that since reaction (B) does not provide new electrons, as would reactions (A) or (C), the resulting value of Townsend's α would necessarily be much lower than in simple gases such as H_2 or Ne.

It is also known that the dielectric strength of the halogen derivatives of methane show progressively falling dielectric strengths as the Cl atoms in CCl_4 are successively replaced by F atoms, and indeed that CF_4 has a dielectric strength relatively little different from that of air. It is suggested that this is due to the increasing importance of reaction (C) compared with reaction (B) as the dielectric strength decreases. The available data (Baker and Tate 1938, Vought 1947) support the above suggestions but, clearly, much more experimental work is required to check the proposals. Various aspects of this work are now being studied in these laboratories.

ACKNOWLEDGMENTS

The authors wish to express their thanks to Dr. F. H. Andrews and many of his colleagues in the I.C.I. Research Laboratories at Runcorn and Widnes for their great kindness in providing samples of special gases and for helpful discussions on certain chemical aspects of the work. They are also grateful to Professor J. M. Meek for his interest in this work, and for many discussions on electrical breakdown in gases, and to Mr. C. A. McDowell for discussions on dissociation processes.

REFERENCES

- BAKER, R. F., and TATE, J. T., 1938, *Phys. Rev.*, **53**, 683.
- CHARLTON, E. E., and COOPER, F. S., 1937, *G.E. Rev.*, **40**, 438.
- CRAGGS, J. D., and MEEK, J. M., 1948, *Proc. Phys. Soc.*, **61**, 327.
- DIBELER, V. H., and MOHLER, F. L., 1948, *Bur. Stand. J. Res., Wash.*, **40**, 25.
- HOCHBERG, B., 1947, *Elektrichestvo*, No. 3, 15.

- HOCHBERG, B., and SANDBERG, E., 1942, *J. Tech. Phys. U.S.S.R.*, **12**, 65; 1946, *C.R. Acad. Sci., U.R.S.S.*, **53**, 511.
- JOLIOT, F. M. FELDENKRAIS, and LAZARD, A., 1936, *C.R. Acad. Sci., Paris*, **202**, 291.
- LOEB, L. B., and MEEK, J. M., 1940, *The Mechanism of the Electric Spark* (Stanford : University Press).
- MASSEY, H. S. W., 1938, *Negative Ions* (Cambridge : University Press).
- MEEK, J. M., 1940, *Phys. Rev.*, **57**, 722.
- MOHR, E. I., and WEISSLER, G. L., 1947, *Phys. Rev.*, **72**, 294.
- POLLOCK, H. C., and COOPER, F. S., 1939, *Phys. Rev.*, **56**, 170.
- RODINE, M. T., and HERB, R. G., 1937, *Phys. Rev.*, **51**, 508.
- STEPHENSON, J. D., 1933, *Proc. Phys. Soc.*, **45**, 20.
- STEVENSON, D. P., and HIPPLE, J. A., 1942, *J. Amer. Chem. Soc.*, **64**, 2766.
- VOUGHT, R. H., 1947, *Phys. Rev.*, **71**, 93.
- WARREN, J. W., and CRAGGS, J. D., 1950 (to be published).
- WEBER, W., 1942, *Arch. Elektrotech.*, **35**, 166.
- WEISSLER, G. L., and MOHR, E. I., 1947, *Phys. Rev.*, **72**, 289.

The Static Coefficient of Friction and the Area of Contact

By R. C. PARKER AND D. HATCH

Messrs. Ferodo Ltd., Chapel-en-le-Frith, Stockport

Communicated by E. N. da C. Andrade; MS. received 11th July 1949

ABSTRACT. An apparatus for measuring the static coefficient of friction has been designed in which the forces tending to disturb the contact area are minimized and which further enables the contact area to be viewed throughout the experiment. Application of a tangential load, less than that required to give limiting friction, has been shown to give rise to a growth of both the real and apparent contact areas and to produce a relative displacement of the surfaces to an extent governed by the tangential load. An examination of the surface topography, by means of an electron microscope, has shown that the earlier type of slip ceases because surface scratches extend the boundaries of real contact. Both the force of adhesion and the tangential force vary linearly with the real and apparent areas of contact. These results are in accord with a cohesive theory of friction, and offer no support for the theory on which friction is attributed to the formation of welded junctions.

§ 1. INTRODUCTION

WHEN attributing the force of sliding friction between solid bodies to the raising of one member over the surface irregularities, or asperities, on the other, Coulomb (1785) discounts the explanation of cohesive force on the grounds that it was not in accordance with his experimental observations that friction is independent of the area of contact. The existence of strong cohesive forces was nevertheless later postulated by Hardy (1936) in his work on static friction, and he found evidence for it in such effects as the doubly refracting nature of pieces torn from a glass slider.

There has been a reluctance to accept a purely cohesive theory of friction, since our knowledge of elastic deformation has been insufficient to explain the observed proportionality between the normal load and the real area of contact. This difficulty remained until Bowden and Tabor (1939) claimed to have shown that the surface irregularities deformed plastically in such a way that the real area of contact was proportional to the normal load. Although Bowden and co-workers recognized that bulk properties of the bodies and their surface dimensions

(Bowden and Hughes 1939) played an important rôle in friction, they nevertheless believed that the major contribution arose from welding of the surfaces at their points of contact (see Bowden and Tabor 1945). An objection has, however, been lodged by Bickerman and Rideal (1939), who showed that it was not, in general, possible to reduce the normal load between two bodies without also decreasing the real area of contact, and, therefore, the force of friction. This, they claimed, was contrary to the conception of plastic flow and cohesion. In reply, Bowden and Leben (1940) pointed out that beyond the boundary of plastic flow there exists an elastic strain which, when released by a reduction in the normal load, would sever many of the points of contact. They also suggested that, as the strain produced by the tangential pull could not be equally shared by all the points of contact, the friction must be determined by the load actually applied during the measurement. Nevertheless, failing further evidence on the relation between force of friction and the real area of contact, Bickerman (1948) and Gemant (1943) still adhere to Coulomb's asperity theory of friction.

In view of the rival claims of the asperity and cohesion theory, it was felt that more experimental evidence was needed to dispel, once and for all, the belief that the asperity theory constituted in itself an adequate explanation for all friction phenomena. The work described in this paper purports to do this, and at the same time provides new information on the dependence of the tangential force of friction on the real area of contact.

For an investigation of this kind elastic materials are unsuitable, since the real area of contact is determined only by the normal load and so does not enable the effects of these two quantities on friction to be separated. There are, fortunately, materials for which change in area of contact may be brought about without change of normal load. For instance, the area of contact between hemispheres made from material of low yield point, such as lead or indium, is a function of the arithmetic sum of the magnitudes of the curvatures of the two surfaces. The area of contact of materials such as lead may also be varied, independent of normal load, by a hysteresis effect which is exhibited on increasing and decreasing the normal load. Preliminary experiments, using the latter phenomenon, showed that the coefficient of friction as normally defined can be increased by 100 per cent, and although these results were only qualitative, they were regarded as sufficiently encouraging to warrant the construction of apparatus for quantitative work.

§ 2. APPARATUS

The design of the apparatus presented the problem of sliding one surface over another without the introduction of any force which might tend to disturb the area of contact. It was also considered desirable to incorporate means for viewing the area of contact throughout the experiment. These requirements demand parallel motion of the sliding surfaces, an arrangement to prevent tilting of the fixed surface, and the use of a transparent material for at least one friction member.

The essential features of the apparatus designed to meet these conditions are shown diagrammatically in Figure 1. Parallel motion of the sliding surfaces is achieved by the use of two glass optical flats, A and B, separated by a thin film of oleic acid. The upper flat, A, was rigidly mounted on an aluminium plate 6 mm. thick, and the lower flat, B, which serves as one friction member, was kept in equilibrium by the second friction member, C, which was thrust upwards by the normal friction force. For the quantitative work described in the first part

of this paper the normal load was maintained sufficient to support the weight of the lower optical flat (40 gm.) and to maintain the thickness of the oleic acid film constant for the duration of the experiment. For the latter work, carried out with an electron microscope, the normal load was increased up to 240 gm. The normal load was applied by means of a string in tension that pulled up a triangular frame, D, whose apex supported the second friction member C (mounted in wax), while to the base of the frame was fitted a rod which rolled in Vee blocks spaced 20 cm. apart. By these means tilting of the friction members was reduced to a minimum. The movement of flat B was effected by light cords attached to the framework E, which was built around the flat in such a manner that the plane of pull was near to the plane of sliding. The cords passed over three pulleys as shown, and were kept in tension by equal weights attached to the free vertical lengths. It was found convenient to apply the frictional forces by means of suspended beakers into which was poured water, accurately measured on completion of the experiment. In

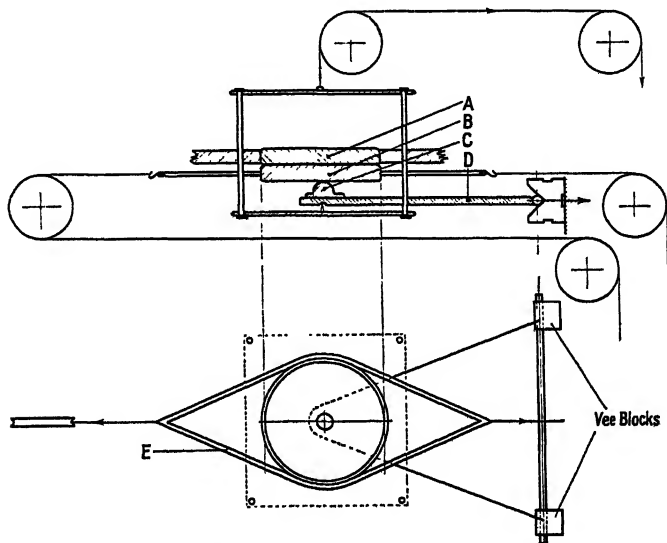


Figure 1. Diagram of apparatus.

some of the experiments it was required to restrict the slip to a pre-selected value, and this was done by limiting the movement of the framework by means of the anvil of a bench micrometer capable of detecting displacements of 0.5 micron. Contact between the framework and the anvil was indicated electrically.

With the above arrangement the area of contact was viewed through a microscope fitted with a $1\frac{1}{2}$ in. objective, and an eyepiece that was fitted with a measuring micrometer calibrated to 2.5 microns. Vertical illumination was used.

To prescribe the area in which contact occurred the friction member, C, was made in the form of a hemisphere. Experiments were carried out with lead and indium of Analar purity. The specimens were made by pouring molten metal from a crucible into hemispherical impressions, which had been formed by pressing ball bearings into a polished $\frac{1}{4}$ in. thick copper plate. The metal poured cleanly, and on solidification its surface was characterized by a regular system of fine cooling ripples (see Graf 1942) similar in appearance to the undulations left on sand by a receding tide. Their distance of separation was approximately 3 microns.

§ 3. EXPERIMENTAL TECHNIQUE

The flat B was first removed, and its lower surface polished with a piece of mutton cloth that had been degreased by extraction with trichlorethylene in a Soxhlet apparatus. The polishing was continued until the contaminating films were sufficiently thin and homogeneous. The latter condition was judged by breathing upon the surface and noting the appearance on the resulting grey film. At this stage a drop of oleic acid was placed on the reverse side and the flat remounted in position. The drop was then spread into a film by rotating and pressing the lower flat upwards, after which it was centralized with respect to the upper flat. Measurements throughout the course of this work showed that the force required to move the flat never exceeded 3 gm. The lead hemispherical member was finally fixed in position, and then slowly moved into contact with the glass.

Initial experiments showed that the apparatus was extremely sensitive to vibration, and vertical oscillations that were transmitted to the friction member gave contact areas far in excess of those required to support the normal load. Indeed, despite the fact that the friction surfaces must have collected a considerable amount of contamination from the atmosphere, values of μ could not be obtained less than unity. The effect of vibration was, however, progressively reduced by partially insulating the apparatus on rubber pads and replacing the light cords by elastic strings. Figure 2* shows a general view of the apparatus.

§ 4. EXPERIMENTAL

(i) *Friction Force and the Area of Contact*

Examination of the contact area, formed by placing the $\frac{3}{8}$ in. diameter hemisphere of lead in position against the glass flat, showed that the resultant distortion did not materially alter the appearance of the cooling ripples or crystal boundaries. In fact these features were still preserved when the hemisphere was distorted to a disc of thickness approximately one-quarter of the original diameter. The preservation of the surface flaws has been attributed by Moorc (1948) to the tips of the asperities undergoing work-hardening and so being harder than the bulk material. This explanation, however, does not seem applicable to this work, since systems of slip bands were observed to traverse the real area of contact and neighbouring valleys without any perceptible change in direction or spacing. The bands merely appeared to be flattened by contact with the glass. Although it was not found possible to obtain a clear photograph of this behaviour, the form of slip bands is illustrated in Figures 3 and 4 (Plate). Figure 3 shows a general view of the contact area, and the slip bands are to be seen as criss-cross lines at the bottom left-hand corner. Other slip bands can be seen at the top right-hand corner. Figure 4 shows an enlargement of slip bands that lie in a valley within the apparent area of contact. In the top left-hand corner all surface features are obliterated by slip, but in the top right-hand corner slip bands can be seen to cross, undeviated, a small plateau. Notwithstanding this preservation of surface identity, the boundary of contact was clearly defined, and was taken as that located by the inner ring of a system of fine interference fringes. On applying a tangential load two unexpected features were observed. The apparent area of contact increased with the tangential load, and this was accompanied by a gradual obliteration in the surface features (see Figure 3). These changes continued until the tangential load was sufficient to cause slip. The term is here used to denote

* For Figures 2-4, 7 and 8 and 11-16, see Plates.

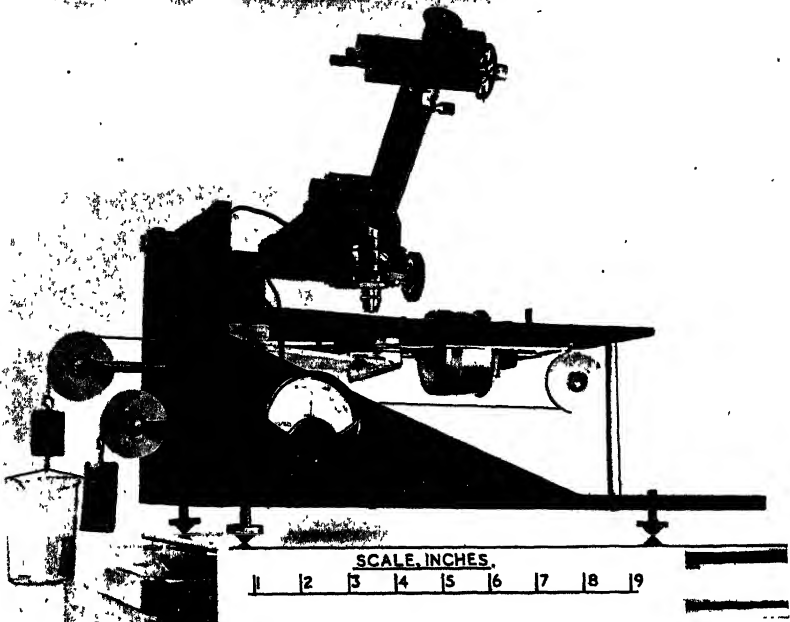
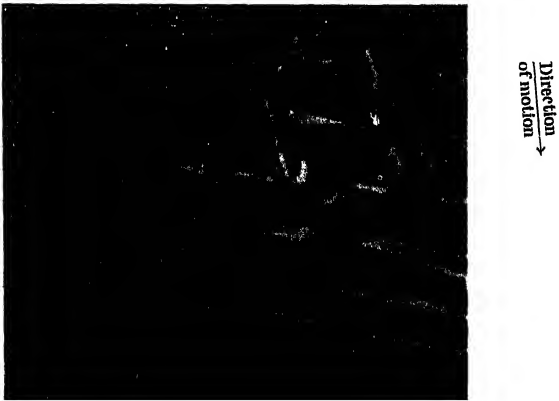


Figure 2. General view of apparatus.



Figure 3. Apparent area of contact between $\frac{1}{8}$ in. diameter lead and glass flat after an interfacial movement of 25.4 microns. Vertical illuminations.



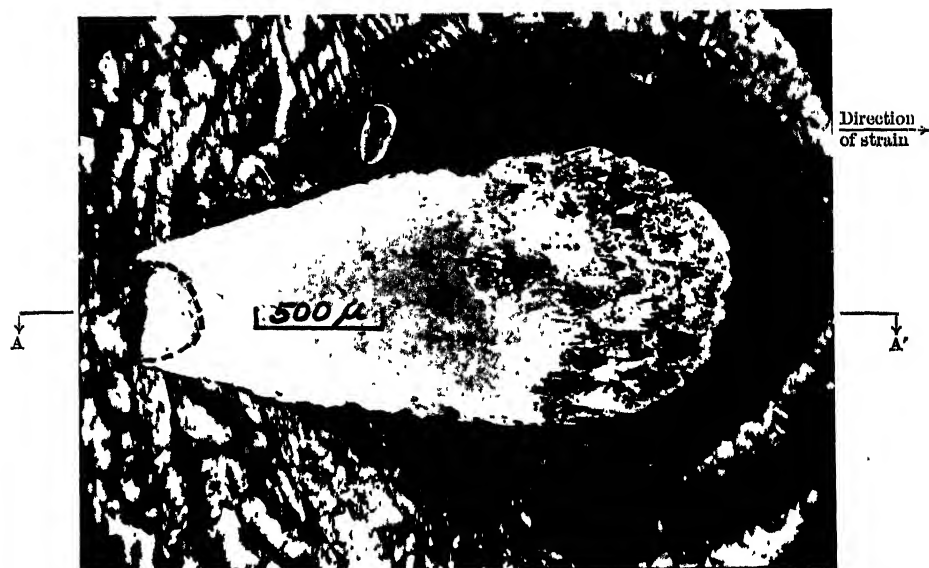


Figure 7. Apparent area of contact between $\frac{1}{8}$ in. diameter indium and glass flat at a tangential load of 2.5 kg. and normal load of 40 gm. Oblique illumination.

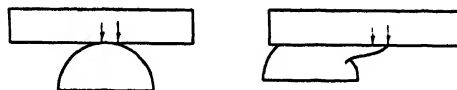
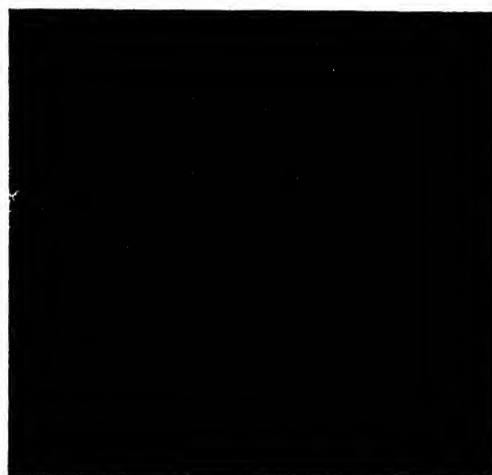


Figure 8. Section through AA', Figure 7, showing the type of distortion.



Figure 11. Lead surface before being placed in contact with the glass.



Direction
of slip
↙

Figure 12 Lead surface after a slip of 6.35 microns. Note scratches on tips of asperities



Direction
of slip
←

Figure 13. Scratches extending to boundary between real and apparent areas of contact. Slip 25.4 microns.



Figure 14. Contact area covered with scratches.



Figure 15. View showing nature of scratches.



Figure 16. Deposit of lead on glass flat after sliding $\frac{1}{8}$ in. diameter lead hemisphere at a normal load of 40 gm.

the point at which the tangential force is sufficient to cause continuous sliding, and in what follows will be referred to as macroslip.

A few measurements of the tangential frictional force for the $\frac{3}{8}$ in. diameter lead specimen showed considerable scatter, and an attempt was therefore made to see whether cleaner conditions might lead to greater experimental consistency. The apparatus was obviously unsuitable for use with physically clean surfaces, since the film of oleic acid was too near the friction surface and could not be easily isolated. Strong (1938), however, has described a method whereby chemical cleaning can be obtained by the use of a blowpipe flame or an electric spark. Accordingly the discharge from an 'Edwards' vacuum leak detector was played over the friction surface for a few minutes at atmospheric pressure, and this produced a surface associated with a very high friction that was sufficiently persistent to withstand exposure to the atmosphere for some hours. The high friction of this cleaned surface was also accompanied by a corresponding increase

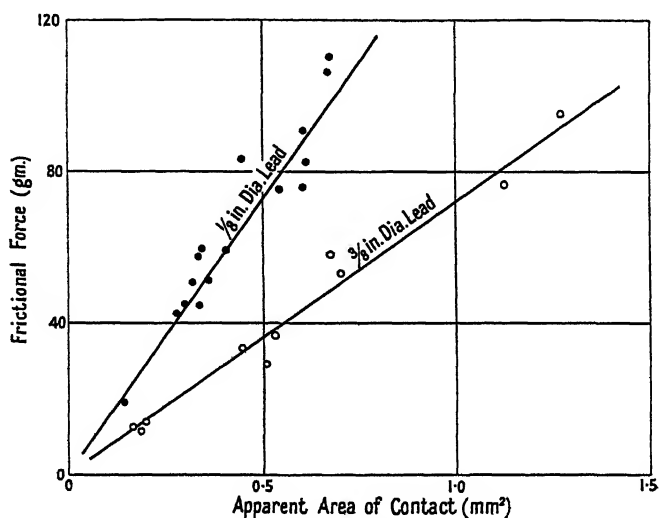


Figure 5.

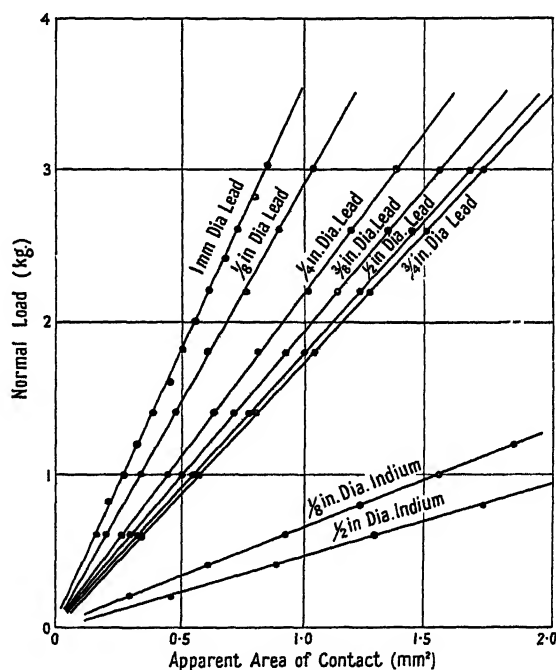
in the area of contact that took place during the increase in the tangential load, and its nature was such that if the experiment was held up at any point prior to slip, the ratio of the area to the applied tangential load was constant.

A variation in the apparent area of contact for the normal load of 40 gm. was obtained by removing the vibration insulation in the apparatus and replacing it in gradual stages. By this means an eightfold increase in area was obtained. A series of measurements was then made of the tangential friction force for the range of contact areas, and the results are shown in Figure 5. It will be seen that there is a linear relationship between tangential force and the apparent area of contact. Results for the $\frac{1}{8}$ in. hemisphere show the most scatter, but even here a correlation coefficient of 0.98 for the above variables showed the relationship to be significant.

It may be recorded, as a matter of interest, that various experimental techniques were tried out in an attempt further to minimize experimental scatter, but in all cases the correlation coefficient remained approximately the same. A part of this scatter may be attributed to any remaining instability of contact, since any

minute displacement of the friction surface would tend to allow the contact area to regain its value associated with the normal load of 40 gm. The variation of friction with crystal direction is also likely to contribute to the scatter, especially as the crystal size of the lead and indium was observed to be of the same order as the contact area. In this connection Gwathmey, Leidheiser and Smith (1948), working with single crystals of copper, have noted a fourfold variation in the static friction with crystal direction.

The relation shown in Figure 5, i.e. that the tangential friction force is proportional to the apparent area of contact, is contrary to Amontons' law and to general experience. If the friction force is attributable to cohesive forces, it is to be expected that the tangential friction force will be proportional to the real area of contact, and for this to be possible the ratio of the apparent to the real area of contact must have been constant. However, any assumption on the nature of



the lead was returned over the same track immediately after macroslip, a sudden increase in the degree of scratching occurred as soon as the lead re-crossed the boundary of original contact area. This increase in scratching was accompanied by an increase in the friction force and the apparent area of contact. In the first traverse of the lead, contaminating films must have been wiped from the original contact area and smeared over the track. The exact nature of this phenomenon depended upon the original state of cleanliness of the glass surface.

(ii) *Adhesion*

The surfaces cleaned by the electrical discharge were characterized by measurable forces of adhesion. With lead the adhesive force was small compared with the normal load, but with indium the adhesive force was so great that sliding produced very great distortion. The indium behaved as a plastic material in that even with a small normal load the area of apparent contact could be increased indefinitely on applying a tangential load. In Figure 7 (Plate) is shown the contact

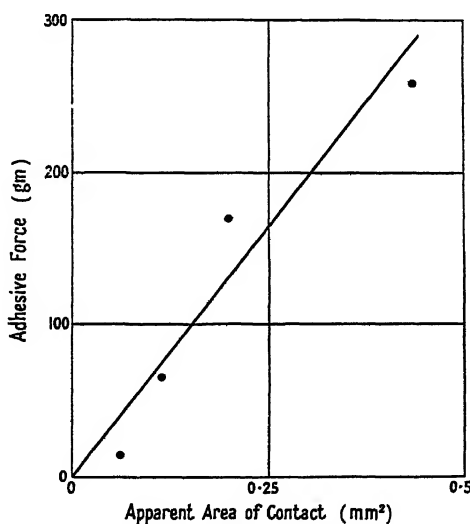


Figure 9.

area formed between a $\frac{1}{8}$ in. diameter indium hemisphere and glass for a tangential load of 2,500 gm. and a normal load of 40 gm. The position of the original area before application of the tangential load is shown outlined. One half only is shown, since the other portion became detached on removing the indium from the glass prior to taking the photograph. In Figure 8 is illustrated a section taken through AA, showing that not only is the shear strength of the material exceeded near the surface, but for some way down. These experiments were satisfactorily repeated in an atmosphere of pure hydrogen, which showed that the wetting of the glass by the indium was not due to the formation of an oxygen bond (King 1942). It may be remarked that previously recorded cases of high adhesion have required either high temperatures (Schnurmann 1941) or outgassing at high temperatures under vacuum (Bowden and Hughes 1939).

The adhesion values of a $\frac{1}{8}$ in. diameter hemisphere of indium were measured against glass, and the results given in Figure 9 show the adhesion force to be linear with the apparent area of contact. In this experiment the glass was not electrically

cleaned, and the area of contact was produced by applying a tangential force with a 40 gm. normal load. Since it is reasonable to expect that adhesion will be proportional to the real area of contact, this relation again confirms the consistency of the ratio of the real to the apparent area of contact.

When Hardy postulated his cohesive mechanisms of friction he pointed out that resistance to displacement of the surfaces along the tangent, but not along the normal, could be expected, since separation along the normal involved the progressive peeling away of the surface. In the case of the compressed hemisphere, any tendency towards recovery is concentrated at the edge of the contact disc over an annulus of a thickness that depends upon the effective range of the surface forces. Although for indium this force of adhesion may be of the same order as the normal load, it does not materially influence the area of contact owing to the short-range nature of these forces. That this is so is demonstrated by the fact that the linear relation shown in Figure 5 is not disturbed by the fact that only certain of the observations were associated with high adhesion values.

(iii) *The Nature of Slip*

To study the growth in the contact areas due to the application of a tangential load it was decided to conduct friction experiments on lead with the tangential load restricted to a small fraction of that required to give macroslip, and to study the appearance of the contact areas on the lead with an electron microscope.

The electron microscopy was carried out with a Metropolitan-Vickers 50 kv. EM.2 tube used in conjunction with a replica technique and gold shadowing. The choice of replica technique was limited owing to the low melting point and softness of the lead, and by the difficulty of dissolving the metal away from such materials as silica. Difficulties also arose from the curvature of the metal surrounding the contact area and the magnitude of the surface irregularities. The former tended to cause bad wrinkles in the film. The technique adopted was to use simple negative replicas formed from a dilute solution of polyvinyl formal and ethylene dichloride. The size of the surface features and the curvature of the hemispheres demanded a half per cent solution which corresponds to a film thickness in the range of 1000 to 2000 Å. The replicas were formed by immersing the hemispheres in the formal solution, after which they were shaken to remove surplus solution and allowed to dry for about five minutes. Two lines were then cut by means of a razor through the replica into the lead in the direction of strain. The replica was removed from the lead under distilled water by folding the film back on itself with the aid of a pair of fine forceps. The surface replicas were then floated, impression side upwards, and set in position on a sample grid. After drying for several hours in a desiccator, the replica was shadowed, at right angles to the direction of frictional strain, with gold atoms at an angle of $\tan^{-1} \frac{1}{4}$ to a calculated depth of 90 Å.

It was planned to begin by obtaining electron micrographs of the lead surface corresponding to an applied tangential load one-tenth of that required to give macroslip. In this experiment, however, it was noticed that even with this applied load some slip took place, and it was therefore decided to study its mechanism in more detail before examining the contact areas. A few subsequent experiments showed that the extent of this slip was determined by the tangential load, and that it was accompanied by a large visual change in the magnitude and

distribution of the real area of contact, the nature of which must have restored equilibrium. This dependence of slip upon the tangential load suggested the feasibility of limiting the slip rather than the tangential load, and the projected experiments were modified in that the bench micrometer was used to increase the slip in steps of 6.35 microns up to a total of 63.5 microns. This latter value was in the neighbourhood of that required to give macroslip. The observed relation between the tangential load and the distance slipped is shown in Figure 10. The point of macroslip, corresponding to static friction as normally measured, is shown by the dotted line, and will be seen to have been approached slowly. This unambiguous determination of the static coefficient of friction is contrary to the view expressed by Bristow (1947).

An examination of the surface of the lead hemisphere throughout these experiments revealed changes in contact areas which, in conjunction with a cohesive theory of friction, satisfactorily accounts for the progressive nature of slip. In the absence of a tangential load the tips of the asperities supported the load elastically, and the electron microscope revealed no evidence of permanent

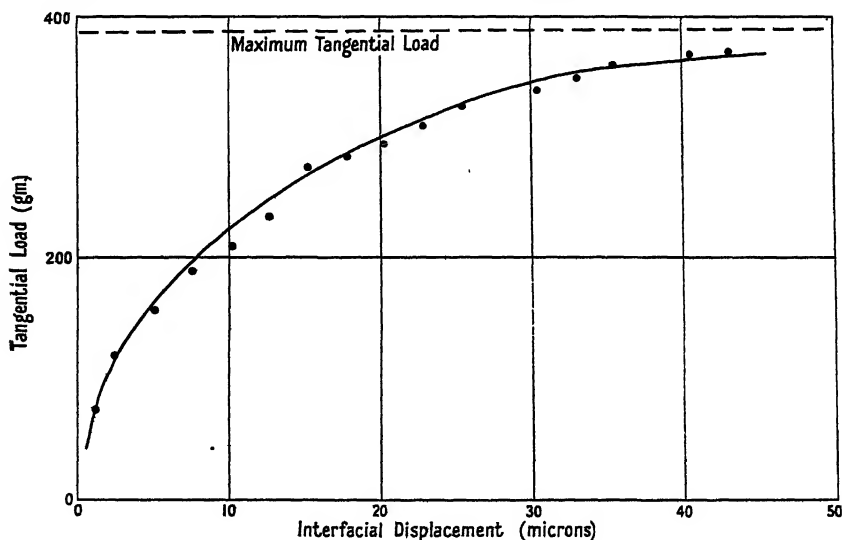


Figure 10.

deformation of the asperities. In fact the shape of the asperities was so well preserved that, on removing the normal load, gold shadowing failed to record any change across the contact area. When, however, a tangential load was applied equal to 10 per cent of that required to give macroslip, the asperities deformed plastically into small plateaux within the apparent contact area. Figure 11 (see Plate) shows an electron micrograph of the lead surface before being placed in contact with the glass, and Figure 12 shows the surface after a slip of 6.35 microns. The tips of the asperities were always scratched, which showed that interfacial movement has occurred, and in Figure 13, corresponding to a slip of 25.4 microns, the scratches are seen to extend up to the boundary dividing the real and apparent areas of contact. With further increase in the tangential load a proportional increase took place both in the slip and in the growth of the real area of contact, and eventually the whole of the apparent area of contact

had contacted the glass and had become covered with scratches (Figure 14). At this point the tangential load reached its maximum value and macroslip ensued. The nature of the scratches is shown in Figure 15 in which they are seen to consist of parallel furrows without any indication of pile-up of displaced material.

§ 5. DISCUSSION

The mechanism of slip has been revealed in greater detail than hitherto, and the sequence of the changes in contact area, and the effect of such changes on the tangential frictional force, can now readily be visualized from the moment of placing the two bodies in contact. When the lead or indium hemisphere is pressed against the flat hard surface bulk plastic deformation occurs, but the surface irregularities are preserved and the tips of the asperities undergo a deformation which is mainly elastic. The real area of contact is proportional to, and increases with, the normal load. Momentary increase of the normal load, such as occurs from vibration, can give rise to a permanent increase in the real area of contact, provided there are no extraneous disturbing forces such as would result from tilting. If to one of the two contacting surfaces a tangential pull be gradually applied, the surface will presumably undergo some elastic deformation due to conservative forces (Rankin 1926), followed by plastic deformation of asperities and an accompanying increase in the real area of contact. These latter two features will occur when the tangential force is only a small fraction of that required to give macroslip. With further increase in the tangential force the original areas at the tip of the asperities become scratched and grow in extent, while new regions of real areas of contact appear. This process of scratching, with increases in the real areas of contact, continues with increase in the tangential load until the whole of the real contact area is covered with scratches, at which point macroslip takes place. If the real areas of contact have been artificially increased by vibration, or by hysteresis occasioned by previous increasing and decreasing of the normal load, the magnitude of the changes will be correspondingly reduced. It is due to these changes in contact area that slip takes place progressively. Thus if a small tangential load less than that required to give macroslip be applied, an out-of-balance system of forces is set up which gives rise to slip. This only ceases when equilibrium is restored by an increase in the cohesive forces that are brought about by a growth in the real area of contact. It is when further growth cannot take place by reason of external geometrical features that slip becomes unlimited.

Throughout all stages up to macroslip the tangential load was found to be proportional to the apparent area of contact, and therefore also proportional to the real area of contact since the ratio of the two areas was deduced to be constant. The concept of frictional force per unit of real area of contact has previously been used by Holm (1938), who termed it the specific force of friction. This manner of expressing the results is obviously more desirable here than the use of the coefficient of friction, for the latter involves the normal load and so introduces confusion.

Observations through the light microscope showed no visible scratching before macroslip, whereas with the electron microscope scratching was apparent throughout. There is no reason to doubt that scratching commences from the moment the conservative forces are exceeded, and this being so, it is difficult to appreciate the distinction made by Schnurmann (1941) and by Bickerman (1948)

between sliding with scratching and sliding without. To restrict Amontons' law, and even the term friction, as these writers suggest, to sliding that occurs without permanent deformation places an unwarranted emphasis on an arbitrary criterion as to whether or not scratching is visible to the naked eye.

An examination of the formation of scratches with the electron microscope shows a number of important features. With the smallest observed slip there were small particles of lead adhering to the glass plate. In illustration of this mechanism, Figure 16 shows lead particles adhering after a slip of 3 mm., while the shadow replica technique from the lead showed that the particles had ploughed parallel-sided furrows in it without any change in the size of the particles. These furrows, moreover, had no lead piled up at their sides, so that the displacement of lead must have occurred by means of transference through the bulk of the material. That this did in fact occur was confirmed by changes in the appearance of the contours of the isolated contact areas through scratching. On the application of a tangential load only one-tenth per cent of that required to give macroslip, areas containing scratches were changed to small plateaux standing in sharp relief to the general background. Figure 17 illustrates a cross section of the apparent contact area and indicates the manner in which the surface topography can be altered by scratching. In (a) the contact angle is much less than $\tan^{-1} \frac{1}{4}$, so that the shadow is indefinite. In (b), however, the metal displaced from A to B increases the contact angle and so gives rise to sharp relief. The growth in the



Figure 17.

(a) Section through contact area before application of tangential load, showing the tip of one asperity in contact with the glass.

(b) After the application of tangential load, showing the volume of metal displaced by scratching to form more contact area.

real area of contact is thus seen to depend upon the depth of the scratch and the angle of contact at the boundary. The fact that scratching is associated with bulk deformation through the material renders any attempt to estimate the quantitative contribution of scratching to the force of friction very difficult. It is to be noted in passing that the ploughing term discussed by Bowden, Moore and Tabor (1943) is not apposite, since their experiments and calculations assumed incorrectly that metal is piled up at the edge of the ploughed furrow.

The preservation of asperities during deformation of relatively hard metals has been described by Moore (1948), although there has been a general belief that plastic deformation of non-conforming materials leads to a coincidence of the real and apparent contact area (see for example Blok (1940)). This belief, here demonstrated to be false, seems to have been held by Bowden and Tabor (1939) in their work on the area of contact between stationary and moving surfaces. These authors measured the area of contact between crossed cylinders by conductance methods, and by microscopic observations of the indentation left after removal of the normal load. The reason they were able to deduce correctly the ratio between the normal load and the real area of contact is shown by this work to be attributed to the constant ratio of the real and apparent area of contact. Their agreement between the electric and visual measurements is explicable in

that the passage of electric currents is not necessarily limited to the restricted areas of close contact but may occur over gaps (Frenkel 1945, 1946). Confirmatory evidence that Bowden and Tabor's crossed cylinders gave rise to apparent contact areas is adduced from the fact that they found it necessary to resort to tapping in order to obtain consistent results. They attributed this effect to the crushing-out of contaminating films, whereas the true explanation would seem to be that the real areas of contact did actually increase. This explanation, moreover, agrees with their observation that reduction in the normal load gave a drop in conductance. Bowden and Leben put forward a tentative theory involving the growth of oxide films.

When the frictional force is measured between an elastically restrained body and a driven surface, oscillations occur, as was observed by Kaidanovsky and Haykin (1933). Bowden and Leben (1940) refer to these as 'stick slips', and attributed them to actual welding of the asperities.

This suggestion of stick slips has resulted in a considerable controversy to which most workers on friction have now contributed. There is a general agreement that stick slips, as a form of relaxation oscillation, would arise given the necessary dynamic properties of the friction-measuring apparatus and the stipulated variation of friction force with sliding speed (see for example Haykin, Lisovsky and Solomonovich 1940, Blok 1940, Morgan, Muskat and Reed 1941, Bristow 1947). Schnurmann and Warlow-Davies (1942) have advanced evidence that the required force-velocity characteristic of sliding would arise from contact electrification. However, no one has yet correlated stick slips observed between a particular pair of materials with the appropriate measurements of the friction force over the velocity range. The concept of limited slip, however, suggests a mechanism whereby stick slips would arise. It is clear that, as the elastic member is deflected, the gradual rise in tangential force will give slip that is counteracted by a growth in the real areas of contact. The elastic member will thus move more slowly than the driven plate up to the point of macroslip. After macroslip the real area of contact is reduced to its initial value, and the phenomenon will be repeated to give a progressive stick slip process. This process incidentally explains the creep of the top slider noted by Bristow (1942). With this explanation there is no need to postulate welding which, after all, is a confusing way of describing intermolecular forces, the nature of which is still largely unknown.

The linear relationship between the force of adhesion and the tangential force of friction on the one hand, and between the real and apparent areas of contact on the other hand, confirms that both forces have a common origin in cohesion. The asperity theory is further disproved in the case of plastic deformation by the fact that these relationships, in so far as the tangential force is concerned, are not affected by the progressive increase in the degree and extent of scratching. These observations have been confined to lead and indium, but must be capable of wider application since even the asperities of hard materials are deformed plastically. Tabor (1949) has calculated that even the minute asperities of tool steel will be deformed beyond their elastic limit under the smallest practical loads.

ACKNOWLEDGMENTS

The authors wish to record their thanks to the Washington Chemical Company for the use of their electron microscope, and to Mr. A. Gard, who kindly assisted with its operation. Special thanks are due to Professor E. N. da C. Andrade

for reading the manuscript and for offering valuable advice as to its presentation. Finally the authors are indebted to Mr. William Smith, Chairman of Messrs. Ferodo Limited, for kind permission to publish this paper.

REFERENCES

- ADAM, N. K., 1938, see *The Physics and Chemistry of Surfaces* (Oxford : Clarendon Press), p. 222.
- BICKERMAN, J. J., 1948, *Lubricating Engng.*, 4, 208.
- BICKERMAN, J. J., and RIDEAL, E. K., 1939, *Phil. Mag.*, 27, 687.
- BLOK, H., 1940, *J. Soc. Automobile Engrs.*, 46, 54.
- BOWDEN, F. P., and HUGHES, T. P., 1939, *Proc. Roy. Soc. A*, 172, 263.
- BOWDEN, F. P., and LEBEN, L., 1940, *Phil. Trans. Roy. Soc. A*, 239, 1.
- BOWDEN, F. P., MOORE, A. J. W., and TABOR, D., 1943, *J. Appl. Phys.*, 14, 80.
- BOWDEN, F. P., and TABOR, D., 1939, *Proc. Roy. Soc. A*, 169, 391 ; 1945, *Annual Reports of the Chemical Society*, 42, 20.
- BRISTOW, J. R., 1942, *Nature, Lond.*, 149, 169 ; 1947, *Proc. Roy. Soc. A*, 189, 88.
- COULOMB, A., 1785, *Mem. de l'Acad. Roy. des Sciences*, p.161.
- FRENKEL, J. I., 1945, *J. Phys. U.S.S.R.*, 9, 489 ; 1946, *J. Exp. Theor. Phys. U.S.S.R.*, 16, 316.
- GEMANT, G., 1943, *J. Appl. Phys.*, 14, 456.
- GRAF, T., 1942, *Z. Elektrochem.*, 48, 181.
- GWATHMEY, A. T., LEIDHEISER, H., and SMITH, G. P., 1948, *Natl. Advisory Comm. Aeronaut.*, Tech. Note No. 1461, p. 37.
- HARDY, W. B., 1936, *Collected Scientific Papers* (Cambridge : University Press), p. 609.
- HAYKIN, S., LISSOVSKY, L., and SOLOMONOVICH, A., 1940, *J. Phys. U.S.S.R.*, 2, 253.
- HOLM, R., 1938, *Wiss. Veröff. Siemens-Werk.*, 17, 38.
- KING, R. M., 1942, *The Glass Industry*, 23, 421.
- KAIDANOVSKY, N., and HAYKIN, S., 1933, *J. Tech. Phys. U.S.S.R.*, 3, 91.
- MOORE, A. J. W., 1948, *Proc. Roy. Soc. A*, 195, 231.
- MORGAN, F., MUSKAT, M., and REED, D. W. J., 1941, *J. Appl. Phys.*, 12, 743.
- RANKIN, J. S., 1926, *Phil. Mag.*, 2, 806 ; 1928/1929, *Trans. Sci. Soc. Roy. Techn. Coll., Glasgow*, 15, 1.
- SCHNURMANN, R., 1941, *Rep. Prog. Phys.*, 8, 71 (London : The Physical Society).
- SCHNURMANN, R., and WARLOW-DAVIES, E., 1942, *Proc. Phys. Soc.*, 54, 14.
- STRONG, J., 1938, *Modern Physical Laboratory Practice* (London : Blackie), p. 165.
- TABOR, D., 1949, *Engineering*, 167, 145 (January).

The Thermal Etching of Single Crystals of Cadmium

BY E. N. DA C. ANDRADE* AND R. F. Y. RANDALL

University College, London

MS. received 25th November 1949

ABSTRACT. The heating of single crystal wires of cadmium in circumstances that permit free evaporation produces a development of crystalline planes. When the basal plane makes a small angle with the axis of the wire there is a characteristic formation of hexagonal pits, the bottoms of which are mirror-like basal planes: when the basal planes make larger angles, characteristic elliptical traces appear, reminiscent of those formed by mechanical extension. The thermal etching reveals, by the formation of steps, thin laminae existing in the unstrained wire: the thickness of these, of the order of 1μ , is the same as the spacing of the glide planes revealed by mechanical extension. The etched figures, in general, indicate that the substructure revealed by straining is already present in the unstrained crystal. The secondary pyramidal faces revealed agree with the calculations of Stranski, Kaischew and Krastanov.

The mirror-like basal planes are sufficiently large to permit the application of an optical method to reveal a scatter in their inclination to the axis. This method shows that there is a variation, round about 20 minutes, of orientation among the basal planes about the direction of growth of the crystals, which were grown by the method of Andrade and Roscoe and showed good single crystal properties by the usual criteria.

§ 1. INTRODUCTION

IN the course of work which we have been carrying out on the effect of surface conditions on the mechanical behaviour of single crystals of cadmium, in the form of 1 mm. diameter wires, it seemed desirable to try to obtain the cleanest possible surface. No form of mechanical polishing is allowable, because the crystal structure of the soft wire is so easily deformed, and our investigations have shown that the contact of electrolytes has a marked effect on the mechanical properties (Andrade and Randall 1948), with the consequence that electrolytic polishing introduces factors not fully elucidated at the time when the work here described was undertaken. It was therefore decided to remove the surface layers of the wire by evaporation *in vacuo*.

In the first case the wire was heated by the passage of an electric current, being contained in a wide tube, the whole walls of which were comparatively cool: later the wire was heated in a wide tube in an electric furnace, from which the ends of the tube protruded. In both cases the effect proved to be not a general evaporation from the surface, but a number of local attacks, which showed interesting features. The heating by passage of a current, either D.C. or A.C., of a wire which had been kept as free from strain as possible resulted in a very marked preferential attack on certain selected glide planes, traces of some of which, but not all, were just visible with special illumination before heating, in spite of the very careful handling. The heating in a wide tube with cooled ends produced etch pits, the flat bottoms of which were hexagonal basal planes with a mirror-like surface. The walls of these pits showed a stepped structure, the 'treads' being close parallel planes, which were likewise hexagonal basal planes. The spacing of these planes was shown to be the same as that of the glide planes revealed by subsequently straining the same crystal. The 'risers' of the steps bounding the pits were $\{10\bar{1}1\}$ planes. The preferential development of the planes in question

may be explained in a general way on the lines first indicated by Kossel (1927, 1928, 1930) and Stranski (1927, 1928, 1929, 1931, 1932).

Single crystals of cadmium were also prepared from polycrystalline wire under conditions that permitted free evaporation during manufacture.

One important general conclusion which this work permits is that there exists in metal single crystals, as normally prepared, a substructure with spacing of the order 1μ which is revealed, but not caused, by strain. This is a point which has already been discussed by Graf (1942), whose paper, published during the war, became accessible to us only after the experimental work was finished.

We have also been able to measure optically the angle between the crystal axes in the basal planes which bound lamellae and to show, as has been argued from x-ray measurements but not directly proved, that there is a scatter, of the order of 1° , in the direction of the axes pertaining to consecutive elements of the substructure.

The experiments here described concern cadmium only. However, thermal etching has revealed a similar substructure in the single crystals of gold and silver which Dr. Cyril Henderson has prepared in this laboratory. An account of the results with these face-centred cubic crystals will be published shortly. The method of thermal evaporation, which is another aspect of what is already known as thermal etching, appears to be a promising one for the investigation of the structure of the real metal single crystal, as distinct from the geometrical ideal.

§ 2. EXPERIMENTAL PROCEDURE

All the crystals used in the earlier part of the investigation were made by the Andrade-Roscoe travelling furnace method (Andrade and Roscoe 1937), in which the wire is enclosed in an evacuated tube whose internal diameter exceeds that of the wire by some 50 per cent. By this arrangement no constraint is imposed on the wire surface, but the curvature of the tube is sufficient to maintain the wire, which has been locally melted, in cylindrical form. No appreciable evaporation takes place, for while the temperature of the metal is near the melting point at the middle of the short furnace, and at room temperature outside the furnace, the solid angle subtended at the centre of the furnace by the space between cold wire and tube, where condensation might take place, is so small that the number of metal atoms travelling in straight lines that can escape from the heated part of the wire is negligible. This is confirmed by the fact that no observable condensation of metal on the walls of the cold part of the tube takes place.

The wires were sealed off in Pyrex tubes at a pressure of about 3×10^{-4} mm. Hg. Before the introduction of the wire the tube had been carefully cleaned and baked out under vacuum conditions at 400°C . and the technique adopted allowed the vacuum to be maintained during the transference of the wire to the clean tube.

(i) *Heating by Passage of Electric Current*

The single crystal wire was placed in the apparatus shown in Figure 1, the transfer being effected with the help of a cradle which was designed to keep the wire strain free. The wire was clamped to brass attachments at A and B; the ends lay in semi-cylindrical grooves in the brass, in which they were retained by small plates, also provided with semi-cylindrical grooves, held to the main attachment by screws. This method of fixing minimized chance strain. The holder A was attached to a terminal block at C, fastened to the brass box MNOP; B was attached by flexible coils of fine wire to E, which was carried by a vertical

steel rod F, rigidly attached to the box MNOP. A current could thus be passed through the wire by terminals attached to C and to the box. The cylindrical prolongation of B passed through a hole in the plate D so that the wire could expand freely when heated by the passage of the current: after the current had been cut off, B could be clamped in position by a screw operated by the head H, which could subsequently be completely withdrawn. The object of this clamping was to

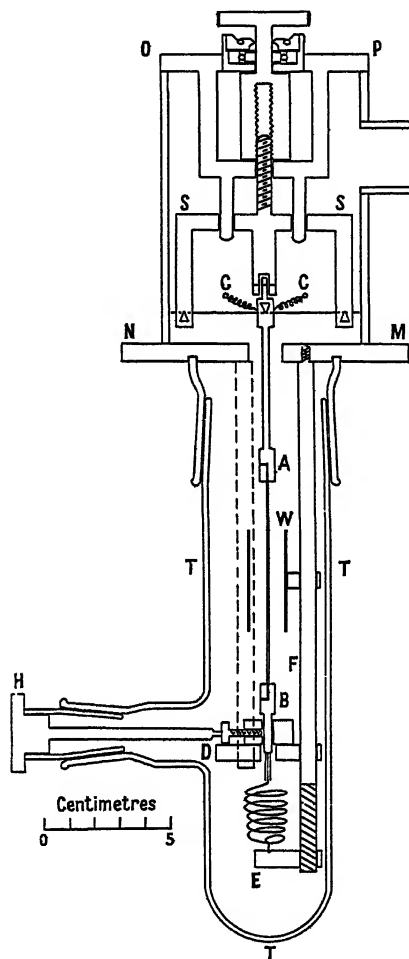


Figure 1.

avoid strain when the wire was being removed. The box MNOP had a removable glass front: the glass tube TTT fitted into a conical holder, from which it could be withdrawn. The arm SS was not involved in the experiments now under consideration. The wire was surrounded by the wide glass tube W; the metallic coating which formed on this tube indicated the amount of evaporation that had taken place.

Experiments were carried out in this apparatus with A.C. of about 4 amp. passing through the wire. A subsidiary experiment in which a 40 s.w.g. thermocouple was attached to the centre section of the wire indicated that the temperature in this region was about 250°C . The weight of metal deposited on the tube W, kindly determined for us by Mr. Tobe, indicated an average

removal of about 30 mg/cm^2 from the surface of the wire, or, say, a layer about $4 \times 10^{-3} \text{ cm.}$ thick. After such thermal treatment the wire showed very markedly intersections of certain of the hexagonal planes with the surface of the wire. The general appearance of such an elliptical marking was reminiscent of the narrow segment that separates two glide packets in a crystal which has been considerably strained, but in the crystal subjected to thermal evaporation the bared portion of the hexagonal plane is considerable, and in consequence the trace shows a deep groove. Figure 2 (Plate I) represents a crystal (for which ψ , the angle made by the basal planes with the wire axis, is 70°) after it has been heated for 15 minutes at about 250°C. by the passage of A.C. This particular crystal had probably been slightly strained in handling or by the weight of the attachment B (Figure 1), which caused faint elliptical traces to appear, but even the faintest of those appearing in the photograph are much more marked than any of those visible before the surface evaporation.

In order to investigate further the development of the elliptical traces we selected a portion of a 25 cm. long crystal made in the normal way, which showed faint elliptical traces under special illumination. We never succeeded in making crystal wires 25 cm. long on parts of which very faint elliptical traces could not be seen, and can only suppose that, although the crystals that we prepare never adhere to the glass, nevertheless, during cooling and consequent shrinkage, the friction of the wire on the glass, which is very small, may be sufficient to produce a few very faint traces of slip bands in the hot wire.

The crystal, for which $\psi = 20^\circ$, was extended by 1% to develop the glide planes slightly and then heated for an hour at 250°C. in an evacuated tube, the ends of which protruded from the furnace, so as to permit condensation. The effect on the traces is shown in Figures 3, 3(a), 4 and 4(a). At A in Figure 3 the very fine line shows the appearance of a slip band before etching. These and other undeveloped traces, which can also be seen in the photograph, require special illumination to make them visible: they are, however, more marked than the very faint traces in the unstrained wire, to which we have referred. In Figure 4 can be seen a developed trace made up of well-defined crystal planes, the broad line representing in a general way the intersection of a basal plane with the wire surface, and the lines AB and BC the development of either pyramidal or prismatic hexagonal faces. No such development has been seen to occur where a faint trace was not first visible with suitable illumination.

With the etching of tungsten (body-centred cubic) crystals, Schmidt (1947), following the work of Johnson (1938), found a difference in the effect of D.C. and A.C. heating, which he attributed to ion migration under the D.C. The effect of D.C. heating was therefore investigated. The general effect was, however, the same as that of A.C. In particular, no difference of effect could be found at the diametrically opposite sides of elliptical traces.

Besides the elliptical traces, representing bared hexagonal planes, many etch pits appeared on the surface. These are discussed in the next section.

(ii) *Furnace Heating*

To avoid, as far as possible, all strain of the crystals it was decided to make the single crystal and subsequently evaporate its surface without any intermediate transfer.

(a) *First method.* A Pyrex tube, whose bore exceeded the diameter of the wire by the usual 50% or so, was ground to form a trough of semicircular cross.

section and a cleaned wire placed in it. The trough was then set horizontally in an evacuated tube 1 cm. in internal diameter, and the small furnace made to travel along it as for normal single crystal manufacture, the difference from the normal case being that the metallic vapour was free to escape into the cool parts of the wide tube. During local melting and recrystallization the evaporation was, in fact, so heavy that the wire was much reduced in diameter. One long single crystal was not formed, but large grains occupying the full diameter of the wire, and some four diameters long, were produced. Such a single crystal grain is represented in Figure 5 (Plate II); it shows sets of facets about 100μ across, traversed by parallel striations indicating lamellae about 1μ across. Such lamellae are shown at higher magnification in Figure 6. Analysis of x-ray back-reflection photographs showed that such grains were single crystals with an orientation spread not exceeding 1° . The striations were traces of $\{0001\}$ planes, as indicated by comparing the results of optical and of x-ray measurements. Optically the reflections showed hexagonal symmetry as the wire was rotated about its axis: they proved to be due to $\{10\bar{1}2\}$ planes. These experiments revealed the existence, in single crystals formed by recrystallization of bulk material, of a lamellar structure similar to that found, for instance, by Straumanis (1931) for zinc crystals formed by condensation of vapour. They did not, however, give an unstrained crystal with cylindrical surface which could subsequently be thermally etched.

(b) *Second method.* To hinder evaporation the semi-cylindrical trough was surrounded by a close-fitting quill Pyrex tube as shown in Figure 7 and the whole

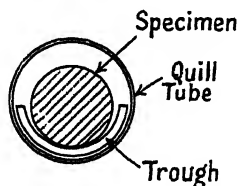


Figure 7.

exposed to the travelling furnace as before. Crystals formed in this way from polycrystalline wire showed, before evaporation treatment, neither facets nor developed elliptical traces. Prominent, however, were pits scattered at random on the surface and, in addition, certain grooves running round the wire in a plane normal to the axis. Surface evaporation was produced by heating the wire, which was protected from chance strain by never being taken from its cradle, at 250°C. , for $\frac{1}{2}$ – $\frac{3}{4}$ hour in an evacuated tube with ends protruding from the furnace. This resulted in an enhancement of the grooves and a development of the pits. The appearance of these grooves was quite independent of the crystal orientation with respect to the wire axis (see Figures 8 and 9 (Plate II), where the close parallel lines are traces of the $\{0001\}$ planes). They were in some cases traced to slight annular marks, visible on the wire as it comes from the drawers, and most probably were due to irregularities in the drawing process. If the intersection of these fine grooves with the wire surface gives edges of very small radius of curvature, then the excess vapour pressure given by the Kelvin equation may account for preferential evaporation in these regions. The grooves would appear to be trivial, and we have not followed the matter up.

The appearance of disconnected pits is characteristic of wires for which the basal planes made only a small angle with the wire axis, i.e. ψ small. When

ψ is 20° or greater, the characteristic development of traces of the basal planes, as exemplified in Figures 2, 3 and 4, takes place.

The etch pits presented different appearances according to their angular position as measured from a plane containing the wire axis and the normal to the hexagonal basal plane of the lattice. The two extremes are represented diagrammatically in Figure 10. In (a) the pit has formed in the neighbourhood of a normal to the basal plane through the wire axis and presents at the bottom a hexagonal plane which gives specular reflection. In (b) the pit has formed in the neighbourhood of a normal to the wire axis which lies in a hexagonal basal plane and shows a series of striations. Actual cases are shown by the photographs of Figures 11 and 12 (Plate II), which represent case (a), and Figure 13, which represents case (b). Should the $\{0001\}$ plane AB, Figure 10(a), make a large angle with the wire axis, the hexagonal plane in which the pit terminates may not be readily visible (Figure 14, Plate III) but becomes so when the wire is extended, with consequent rotation of the $\{0001\}$ planes. Figure 15 shows the pit of Figure 14 after an extension of 250%. The mirror-like plane that terminates the pits was identified as an $\{0001\}$ plane by x-ray back-reflection photographs. The bright lines in Figure 12 are given by reflection from the small portions of

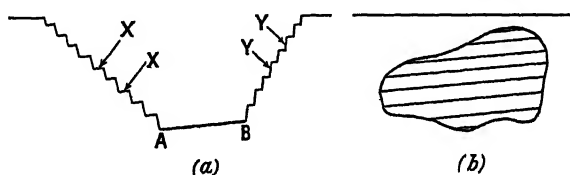


Figure 10.

hexagonal basal planes which form steps in the sides of the pits, e.g. X in Figure 10. Similar reflections may be obtained from the small portions of planes Y in Figure 10 in pits with one or more incomplete sides. The angle between the hexagonal basal plane and the planes Y forming the sides of the pits was found by means of an optical goniometer to be 65° . This corresponds to the angle to be expected if the Y planes are $\{10\bar{1}1\}$ pyramidal planes, the directions of the edges of the bared hexagonal basal plane ensuring that the planes are $\{10\bar{1}k\}$.

It is clearly of interest to know if the laminae revealed by thermal etching correspond to those which appear as glide packets when the wire is strained. Observation under the microscope made it clear that there was a general correspondence, in so far as certain of the planes revealed by etching could be seen to develop further during strain. As a quantitative check the average thickness, normal to the basal plane, of some 100 lamellae was measured as they appeared after an extension of 5%: the figures in the two cases were 0.75μ and 0.85μ respectively, which, considering that a certain arbitrariness is introduced in determining whether certain faint marks are to be counted or not, is satisfactory. With large strains of 200% or so fresh slip bands appear, with a consequent diminution of average thickness of the laminae to about $4/5$ of that indicated by the etch figures. Hence there appears little doubt that the defects of lattice which lead to the laminar appearance of the crystal where the surface has been removed by evaporation are the same defects that cause the glide packets revealed by strain.

(iii) *Lack of Parallelism of Packets*

The slabs of crystal lying between two consecutive 'preferred' planes, that is, planes revealed as privileged either by thermal etching or by the glide produced by stress, will be called a packet. Now if the preferred planes, or transition layers, as W. G. Burgers (1947) calls them, contain a set of parallel dislocation lines of the same sign (see Burgers 1947, Figure 4) there will be a small angle between the axes of successive packets, which will be evidenced by a lack of parallelism between consecutive basal planes revealed by thermal evaporation.

Experiments were therefore carried out to measure by optical reflection variations of angle between near planes revealed by thermal evaporation. The optical arrangement adopted is shown in Figure 16. The lamp filament A is focused on a 0.1 mm. vertical slit and an image of the slit, magnified 3 times, is formed, after reflection from the mirror surface E, on the front of the objective F of a microscope.

The objective is half covered by the stop G, the vertical edge of which is parallel to the image of the slit. The microscope is focused on the face of E. Rotation of E in an anti-clockwise direction causes the image of this face to disappear

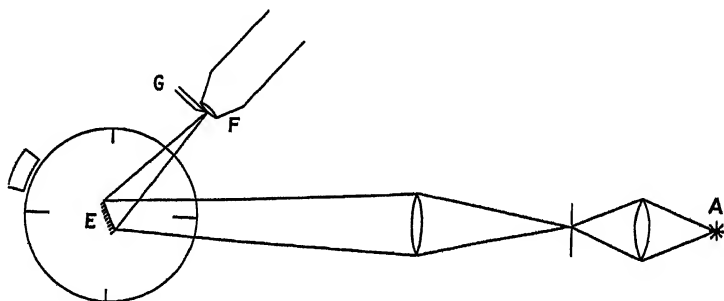


Figure 16.

sharply when the image of the slit crosses the edge of the stop. If in place of the mirror E, used for test, a crystal with hexagonal planes exposed by etching be used, observation on the extinction of various facets constitutes a test for lack of parallelism.

The crystal was mounted on a goniometer (which allowed angles to be read to within 1') with the portion under observation on the axis of rotation. It was prepared and transferred to the goniometer with great care, to reduce the possibility of chance strain to a minimum; the fact that the changes of angles did not progress systematically along the wire suggests that the wires were, in fact, undistorted. As a test, some measurements were carried out with a mirror surface on which all but certain strips, of about the same width as the exposed crystal faces, were rendered non-reflecting. These measurements showed that the method was satisfactory.

The hexagonal planes exposed by evaporation at the bottom of the pits gave sharp reflection and appeared, as far as these experiments can indicate, to be optically flat. Figure 17 (Plate III) shows a typical pit, and Figure 18 one corner under higher magnification. Owing to diffraction the steps at the side of the pit are never sharply defined.



Figure 2. $\times 20$. $\psi = 70^\circ$.
 $\frac{1}{4}$ hr. at 250°C .



Figure 3. $\times 100$. $\psi = 20^\circ$.
1 hr. at 250°C .



Figure 4. $\times 100$. $\psi = 20^\circ$.
1 hr. at 250°C .



Figure 3 (a). $\times 300$.



Figure 4 (a). $\times 300$.

Thermally etched 1 mm. cadmium wires.

PLATE I.

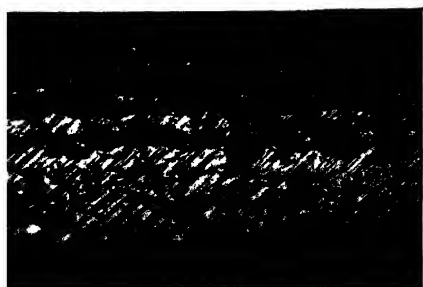


Figure 5. $\times 40$.



Figure 6. $\times 600$.

Crystal grain; $\psi = 45^\circ$.



Figure 8. $\times 20$.



Figure 9. $\times 150$.

$\psi = 60^\circ$; showing annular marks.



Figure 11. $\times 100$. $\psi = 4^\circ$.
See Figure 10 (a).

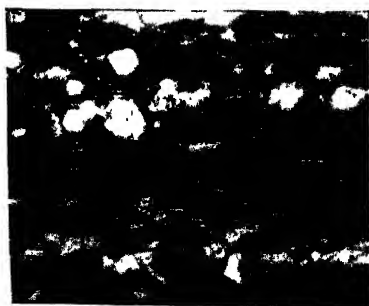


Figure 12. $\times 100$. $\psi = 2^\circ$.
See Figure 10 (a).



Figure 13. $\times 60$. $\psi = 10^\circ$.
See Figure 10 (b).

Etch formations in 1 mm. cadmium wires.

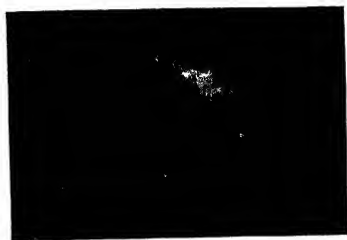


Figure 14. $\times 80$. $\psi = 30^\circ$.
No extension.



Figure 15. $\times 60$.
As Figure 14, but $2\frac{1}{2}\times$ extension.



Figure 17. $\times 1,200$. $\psi = 4^\circ$.



Figure 18. $\times 1,800$. $\psi = 4^\circ$.
Etch pit formations in 1 mm. cadmium wires.

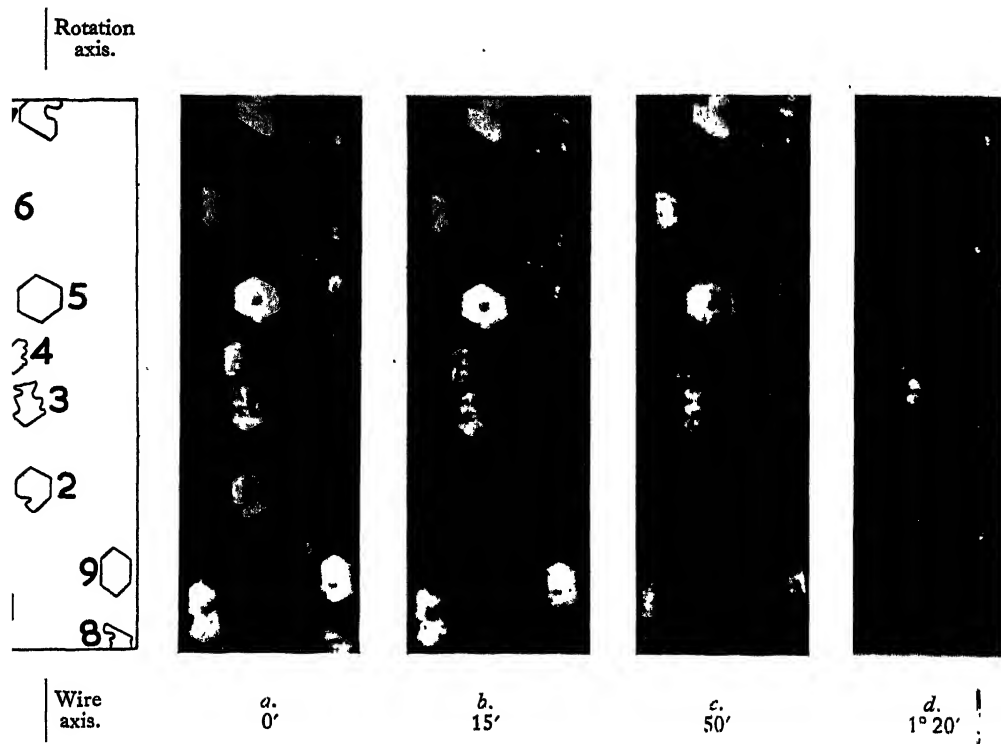


Figure 19. $\times 70$. $\psi = 4^\circ$.

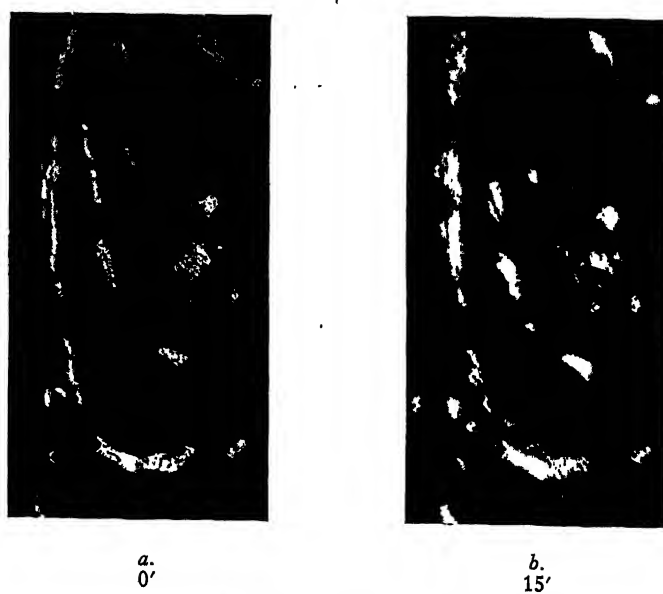


Figure 20. $\times 100$. $\psi = 17^\circ$.

Rotation of thermally etched 1 mm. cadmium wires to investigate parallelism of facets.

With a wire for which the angle ψ made by the hexagonal planes with the wire axis was 4° , mounted with the wire axis vertical (position 1), the appearance of a typical group of pits is shown in Figure 19(a), Plate IV. Figures 19(b), (c) and (d) were taken after rotations from (a) of $15'$, $50'$ and $1^\circ 20'$ respectively. In (b) pits 2 and 8 have already disappeared; in (c), in addition, 1, 4 and 9 are going; in (d) all have vanished except a part of 3. In all cases the edges remain after the corresponding main reflection plane has gone, probably owing to diffraction. In estimating the angle at which extinctions took place allowance was made for difference of position with respect to the axis of rotation. In the case of visual observations it was found that the position of extinction of a given facet could be reset to better than $\pm 5'$. In the case of photographic treatment the pictures were taken at about $15'$ intervals.

Figure 20 shows another crystal, with $\psi = 17^\circ$, in two positions differing by $15'$. It will be seen that in (b) a crystal surface just below the mid-point has disappeared, and close examination will reveal other differences.

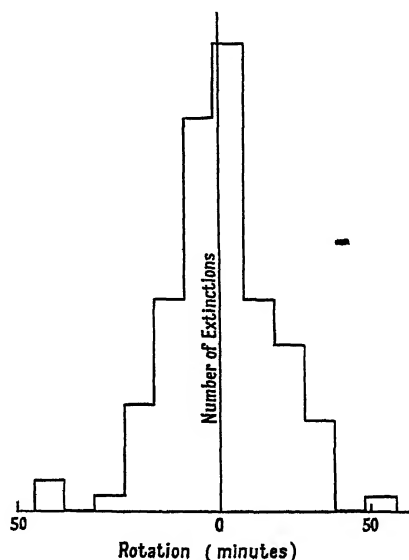


Figure 21.

Six crystal wires were tested with the surfaces of the basal planes on or near the axis of rotation and the wire axis in the same vertical plane as the axis of rotation. Further, one of the crystals was tested with the wire axis normal to the axis of rotation so as to test for lack of parallelism in the direction of growth. In this position eighteen planes were measured optically and no difference of angle greater than $5'$ could be detected.

For the six crystals investigated with the axis of rotation near the wire axis the variation in angle between the facets was determined relative to an arbitrary mean value. This gave in all 113 deviations from the mean, for which a frequency histogram is plotted in Figure 21. The mean deviation is $18'$, which expresses the lack of parallelism.

Our conclusion thus is that, as regards potential glide planes, there is a probable variation of angle about the direction of growth (projection of wire axis on glide

planes) of about $\pm 20'$. On the other hand, there is no variation of angle within the limit of accuracy of our observations (about $\pm 5'$) about a line at right angles to the direction of growth.

It should be emphasized that these results are presented as preliminary, the normal distance between the planes concerned not having been measured, and the number of measurements not being sufficient for a proper statistical handling. The results suffice, however, to illustrate the method.

§ 3. DISCUSSION

The phenomenon of thermal etching in single crystals has been observed with tungsten filaments by Johnson (1938) and by Smithells (1936) who has drawn attention to the structural features revealed by the differential rates of evaporation from different crystal faces. Smithells has published photographs showing the development of large hexagonal surfaces and of planes very regularly spaced at distances of the order of 1μ . Johnson also records the exposure of smooth $\{100\}$ faces by evaporation and several interesting observations on the surface migration of ions with D.C. heating, but failed, apparently under the conditions of his experiments, to reveal any block structure of the order of 1μ . Neither of these authors make any reference to the possible relation of these surfaces exposed by evaporation to the surfaces bared by glide under mechanical forces. Which are the glide planes of tungsten at high temperatures does not seem to have been determined, but from Andrade and Chow (1940) they would be anticipated to be $\{123\}$. Smithells does not state the designations of the close planes revealed in his photographs. In discussing block structure, Graf (1942) describes the systems of parallel markings left on the surfaces of cast metals which have been carefully cooled.

The structure of small cadmium crystals formed from the melt or from the vapour has been described and discussed by Straumanis and by Stranski and his collaborators. Straumanis (1931) has shown that crystals of hexagonal plates, of thickness about 1μ , laid down in steps, the surfaces of the steps being $\{0001\}$ planes. The most frequent of the inclined faces were $\{2025\}$ followed closely by $\{10\bar{1}2\}$ and $\{2023\}$. He also observed the effect of vaporization on the small crystals of a deposit formed on a glass rod, and likewise observed the baring of $\{0001\}$ surfaces. He produced glide lamellae in a small zinc crystal about 1 mm. high and less in diameter by pushing it with the point of a needle, and found a thickness of 0.8μ (1932).

Except, then, for the work on tungsten single crystals, there do not appear to be any observations on the effect of evaporation on single crystals, and all that was clearly established with tungsten was the existence of steps due to the baring of unspecified planes which were separated by, as far as can be judged by the photographs, about 1μ .

The first thing that is clear from the present observations is that on the surface of a single crystal, prepared as carefully as possible and never removed from its cradle, there are spots of high free energy, where the vapour pressure is abnormally high. From these, pits form which show the preferential development of the basal planes and the secondary pyramidal faces explained by Kossel and Stranski on considerations of the number of near neighbours of different classes among the ordered atoms. The pyramidal faces found by us from a consideration of

x-ray and optical methods were $\{10\bar{1}1\}$, which agrees with the calculations of Stranski, Kaischew and Krastanov (1934) for the close-packed hexagonal lattice from which cadmium deviates but little.

If the basal plane makes a very small angle with the wire axis the etch pits appear to form in lines parallel to the wire axis. This may be due to drawing marks which are still visible after the crystal has been made. High vapour pressure at the sharply curved edges of the marks would then initiate evaporation. When the basal planes are nearly parallel to the wire axis, all directions in the surface through a given point are nearly equally favoured as regards removal, and there seems to be no reason why a pit should develop in one direction rather than another, which accounts for the symmetry of the pits about a centre. When ψ is larger the line of intersection of the basal plane with the surface is, of course, in the basal plane, while the normals to this line which lie in the surface make a considerable angle with the basal planes. Hence removal is easier along the line of intersection than along a normal to the surface, which accounts for the development of the 'glide surfaces'.

For larger angles the etch pits tend to extend along the intersection of one particular basal plane and give rise to elliptical traces. An intermediate effect is shown in Figure 4, where the attack has bared pyramidal faces and so led to a hexagonal outline.

All our experiments show the basal planes as the most stable. The evaporation of prepared crystals showed $\{10\bar{1}1\}$ to be the most stable pyramidal plane, in agreement with Stranski, Kaischew and Krastanov. In the case of the crystal evaporated during preparation the $\{10\bar{1}2\}$ planes appeared. The various pyramidal planes have nearly the same binding energy and have all been detected by different observers with cadmium crystals prepared by different methods.

The spacing of the $\{0001\}$ planes is of the order 1μ , and this is in accordance with the observations of Graf and appears to be the unit of the fundamental secondary structure.

Thus a significant conclusion to be drawn from the work is that the glide planes are apparently no more than developments of a structure already present in the crystal. Photographic observation of the extension of an etched crystal may be expected to confirm the identity of the etch lines and glide planes.

The variation of orientation recorded between different glide planes is almost certainly about the direction of growth and not about the digonal axes, as there are three of these and there is no apparent reason why any one should take preference over the other two. So far shear has always been near the direction of growth, and experiments in which this is not so should distinguish between the digonal axes and the direction of growth as directions about which the departure from parallelism occurs.

The method of thermal etching thus has considerable potentialities in the investigation of the secondary structure of single crystals, and should be able to show conclusively that the occurrence of preferential glide on certain selected planes is dependent on a structure already present.

ACKNOWLEDGMENTS

Our best thanks are due to Mr. R. King for helpful discussions on certain points. The junior of us has been supported by a grant from the Ministry of Supply, which is gratefully acknowledged.

REFERENCES

- ANDRADE, E. N. DA C., and CHOW, Y. S., 1940, *Proc. Roy. Soc. A*, **175**, 290.
 ANDRADE, E. N. DA C., and RANDALL, R. F. Y., 1948, *Nature, Lond.*, **162**, 890.
 ANDRADE, E. N. DA C., and ROSCOE, R., 1937, *Proc. Phys. Soc.*, **49**, 152.
 BURGERS, W. G., 1947, *Proc. K. Ned. Akad. Wet.*, **50**, 595.
 GRAF, L., 1942, *Z. Elektrochem.*, **48**, 181.
 JOHNSON, R. P., 1938, *Phys. Rev.*, **54**, 459.
 KOSSEL, W., 1927, *Machr. Götting. Ges.*, **S.**, 135; 1928, *Leipziger Vorträge, S.*, **1**; 1930, *Naturwissenschaften*, **18**, 901.
 SCHMIDT, R. W., 1942, *Z. Phys.*, **120**, 69.
 SMITHELLS, C. J., 1936, *Tungsten* (London: Chapman and Hall).
 STRANSKI, I. N., 1927, *Jahrb. Univ. Sofia*, **24II**, 297; 1928, *Z. phys. Chem.*, **136**, 259;
 1929, *Ibid.*, **A 142**, 453; 1931, *Ibid.*, **B 11**, 342; 1932, *Ibid.*, **B 17**, 127.
 STRANSKI, I. N., KAISCHEW, R., and KRASTANOV, L., 1934, *Z. Kristallogr.*, **88**, 326.
 STRAUMANIS, M., 1931, *Z. phys. Chem.*, **B 13**, 316; 1932, *Z. Kristallogr.*, **83**, 29.

The Lattice-Parameters of Clear Crystalline Quartz

By H. D. KEITH

H. H. Wills Physical Laboratory, University of Bristol

Communicated by N. F. Mott; MS. received 30th June 1949

ABSTRACT. Some experimental results are presented showing that the calibration of Debye-Scherrer x-ray diffraction cameras by means of clear crystalline quartz may lead to appreciable errors unless the parameters of the quartz are first determined by a direct method. The parameters of a sample of clear Brazilian quartz have been found to be in serious disagreement with the accepted values, and variations in the parameters of synthetic quartz, with an apparent dependence on growth temperature, have also been observed. Taken in conjunction with previous results of other workers, the experiments show that quartz parameters vary from sample to sample by amounts as large as 0.01% according to the impurity content. The previously proposed quartz calibration standard is rejected.

§ 1. INTRODUCTION

THE investigations to be described were carried out with a 'Unicam' cylindrical x-ray powder camera of 19 cm. diameter. With a camera of this type it is desirable to calibrate the fiduciary knife-edges directly, but an alternative method has been widely used in which standard values of the quartz parameters are employed to determine the camera angle by extrapolation (Bradley and Jay 1933, Lipson and Wilson 1941a). In some focusing cameras a direct calibration is either impossible or extremely difficult, and a secondary standard must then be employed.

For the standard Bragg angles of quartz, Bradley and Jay suggested those calculated from Bergqvist's value of the a parameter and the assumed value, 1.10002, for the axial ratio at 18°C. Using a camera of 9 cm. diameter, they showed that this value was an improvement on an earlier figure due to Groth. Both the a and c parameters were redetermined with another sample of quartz by Lipson and Wilson (1941a), and these authors ensured a much higher accuracy by making a large number of measurements with three different characteristic radiations.

Their parameters, obtained with a 19 cm. camera, are approximately 0.007% larger than those of Bradley and Jay. In revising the quartz standard, however, Lipson and Wilson apparently studied one sample of quartz only, possibly relying on the fact that Bradley and Jay had previously found no significant differences in the parameters of two samples of clear quartz of different origin. In view of the method adopted by Bradley and Jay, and the fact that they used only two samples, it is difficult to avoid the conclusion that they did not adequately establish the invariability of the parameters of clear crystalline quartz of different origins.

Both the quartz and the direct methods were employed for the calibration of the camera used in our investigations, and good agreement was found between the directly determined camera angle and that obtained from quartz angle extrapolations if the quartz Bragg angles of Bradley and Jay were employed. When the angles given by Lipson and Wilson were used, however, the agreement was not as satisfactory, a result which had previously been observed in this laboratory by Mr. P. Vousden when he calibrated a separate, but similar, camera. In all these calibrations the specimens were prepared from clear Brazilian quartz derived from the same large single crystal. On account of the agreement between these independent observations and of the importance of accurate camera calibration, the parameters of this quartz were determined directly. The relative accuracy of the parameter measurements is of the order of 1 part in 10^5 , and their absolute accuracy must approach this very closely.

§ 2. DISCUSSION OF THE EXPERIMENTAL TECHNIQUE

(a) *Direct Calibration of the Camera*

Two methods described by Lipson and Wilson (1914a) were employed for the direct calibration. With a spectrometer table of large radius the camera angle was found to be 85.299° ; by measuring the radius of the camera and the chord width of the knife-edges it was calculated to be 85.300° . The goniometrical method is more reliable and eight independent measurements of each knife-edge were made, giving a spread of 0.0065° in the camera angle and a probable error of $\pm 0.002^\circ$. In all the later determinations of the parameters of quartz the value $\theta_k = 85.299^\circ$ has been used.

(b) *Temperature Control*

To provide the necessary constant temperature conditions for the experiments, an air thermostat was built which would house the entire camera assembly, and this maintained the temperature constant to within $\pm 0.2^\circ\text{C}$. for many days. This method is valid, however, only if the specimen is not heated appreciably by the x-ray beam during the exposure. Both energy calculations and confirmatory experiments show that this does not happen.

(c) *Discussion of the Inherent Errors*

The geometrical and absorption errors in non-focusing powder cameras can be eliminated readily by extrapolation (Bradley and Jay 1932, Lipson and Wilson 1941 b, Taylor and Sinclair 1945, Nelson and Riley 1945), and the limit of accuracy which can be attained is determined almost entirely by the reliability of the film measurements. Insufficient attention has been paid to the accuracy limits

imposed by visual setting errors, and consideration of these factors has led us to modify the existing technique in the following respects. We prefer (i) higher microscope magnifications, and (ii) somewhat thinner specimens than have previously been considered profitable.

With regard to (i) above, Lipson and Wilson recommended a microscope magnification of 3 diameters for all the film measurements, but we consider that a larger magnification is advantageous. At 3 diameters it is difficult for an observer to distinguish with ease and certainty the true density maximum of a diffraction line even though his settings may have a good reproducibility. Errors may also be encountered in setting on the knife-edge shadows where there is a marked contrast in brightness between the two halves of the field of view. Such errors were, in fact, found in some of our early experiments, and were investigated in the following way.

A scratch made with a razor-blade or a fine diffraction line in the neighbourhood of $\theta = 45^\circ$ was taken on the film and the distance separating it from a knife-edge shadow measured. The film was then reversed in direction and this distance re-measured. Any discrepancy in the two results suggests an asymmetry in the illumination which should be avoided. Reliable settings can be obtained if the contrast at the edge of the shadow is reduced by placing parallel to it a sharp straight edge leaving only a very narrow slit of bright illumination. The difference between the length of the film as measured with and without this contrast reduction is the sum of the errors at the two edges in a typical film measurement. With a magnification of 3 diameters the magnitude of these errors is roughly a length of 0.002 cm. on the film at each edge, corresponding to 0.002% in the measured parameters. If higher magnifications are used, this source of error is no longer important since it then represents a correspondingly shorter length on the film.

It is important to use high magnification in the direct calibration of the knife-edges (§ 2 (a)) because a deviation of 0.002 cm. in the measured position of each edge corresponds to a change of 0.0065° in camera angle. This produces, in turn, errors of approximately 0.0025% in all subsequently measured parameters and also a tendency to curvature in the extrapolation plots.

All the film measurements on which our parameters are based have been made with a Cambridge Universal Measuring Machine, using 3 diameters magnification in some cases and 15 diameters in others. The Bragg angles measured with different magnifications on the same film are in good agreement, allowing for experimental error, but the spread in the experimental values is significantly less in the case of the higher magnification.

Consider now (ii) above. Very thin specimens have been used for the precision parameter determinations because of the consequent reduction in the breadth of the diffraction lines, which allows more precise measurements to be made. The spread of the experimental points on an extrapolation plot is thereby diminished, and the number of parameter determinations necessary to establish a given precision in the averaged result is reduced. This lowers the required number of laborious least squares calculations (see § 2 (d)), and the slightly increased exposure times required are no serious disadvantage. In all these experiments the specimens used were between 0.2 mm. and 0.3 mm. in diameter compared with diameters ranging from 0.5 mm. and 1.5 mm. used by previous workers.

The dependence on line breadth of the accuracy of each parameter measured was established by the following argument. Rymer and Butler (1945) showed that the standard deviation σ of visual settings on a broad diffraction line is closely proportional to the square root of its half-intensity breadth B . The relative error of setting r can thus be written in terms of B as

$$r = \text{const. } \sigma/B = \text{const. } B^{-1/2}. \quad \dots\dots(1)$$

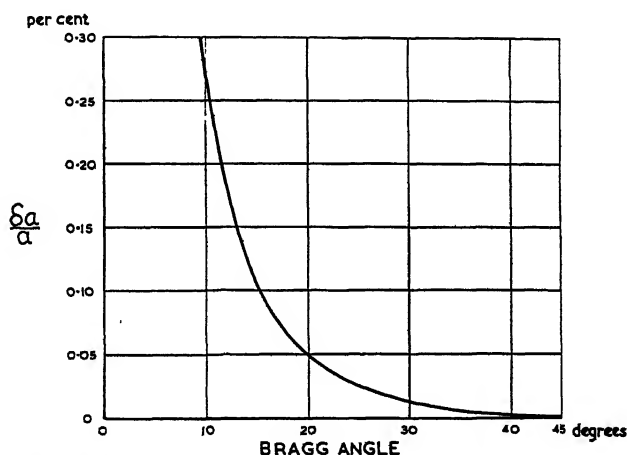
Inserting this into the equation

$$|\delta a/a| = |(rB/2R) \cot \theta|, \quad \dots\dots(2)$$

given by Ekstein and Siegel (1949), for the corresponding error in the parameter we obtain the result

$$|\delta a/a| = \text{const. } B^{1/2}. \quad \dots\dots(3)$$

If an extrapolation involving reflections over a wide range of Bragg angles is to be used care must be taken to keep μr , the product of the linear absorption



Relative error $\delta a/a$ in parameter as a function of Bragg angle.

coefficient and radius of the specimen, greater than 1.0 since otherwise the appropriate extrapolation against $\frac{1}{2}(\cos^2 \theta/\theta + \cos^2 \theta/\sin \theta)$ will not be completely linear. When a least squares calculation is employed with such an extrapolation, using Nelson and Riley's table of this trigonometrical function, it is advisable to correct the points corresponding to Bragg angles less than 45° for the geometrical error due to the finite length of the specimen irradiated (Bradley and Jay 1932, Lipson and Wilson 1941 b). The author has re-calculated the magnitude of this error and obtains a focusing term $\{1 + (R/AX) \cos 2\theta\}^2$ in place of that given by Lipson and Wilson. The relative error in parameter as a function of θ for values less than 45° , calculated from the assumption that $R=95$ mm. and $AX=150$ mm. and from the vertical slit lengths given by Bradley, Lipson and Petch (1941), is plotted in the Figure; the use of this will allow the necessary reduction in the parameters to be readily made. Although extremely fine lines could be obtained with vertical x-ray tube foci, this error prevents their use for these experiments since little focusing is then present in the vertical plane and the elimination of the error by extrapolation would be too troublesome.

(d) Extrapolations

Graphical extrapolation methods can only be applied directly to the simplest structures, and even then do not readily allow the attainment of the high accuracy of which the powder method is capable. Favourable conditions for deciding the best line through the experimental points are usually incompatible with the use of a large scale of ordinates to give high accuracy in reading the extrapolated result. The method of least squares for calculating the most probable result from the experimental points (Cohen 1935) is equally applicable to any structure. It suffers, however, from one serious disadvantage in that it does not, of itself, make any allowance for widely scattered experimental values due either to computational mistakes or to setting errors. This is done automatically by inspection on a graphical plot. The difficulty can be obviated in the case of cubic structures by drawing a rough plot and assigning weights to the points according to their agreement with the extrapolation line. For more complicated structures, such as quartz, alternative methods can be employed.

In our experiments each powder photograph for a given specimen was measured several times and the Bragg angles averaged. Several photographs of each specimen were taken, measured and the average also taken over these results. This procedure has been used throughout to reduce the computational labour.

Before normalizing the equations obtained from the accepted values of the Bragg angles each was weighted in accordance with the following criteria:

(i) All the equations were first multiplied by $\tan \theta$. It is important not to omit this since otherwise we attribute equal value to the measurements at all Bragg angles. The sensitivity of the parameters given by $\delta a/a = \cot \theta \delta \theta$ is such, however, that a specified error in a Bragg angle is more serious in its consequences the smaller that angle.

(ii) Further subsidiary weighting factors were then applied to a small number of the equations to allow for the inevitably reduced accuracy of measurement in the case of fainter lines and of overlapping lines.

The parameters were calculated separately for the α_1 and α_2 lines and the results averaged with the respective weights of 2 and 1. The results of these separate calculations were in agreement to slightly better than 1 part in 10^5 except in the case of one synthetic sample where agreement to 1 part in 8×10^4 only was obtained. Finally, a correction was applied for the refractive index of the quartz.

§ 3. EXPERIMENTAL RESULTS

All the powder photographs were taken at 25° c. with copper radiation from the horizontal focus (8 mm. \times 1 mm.) of a Metropolitan Vickers 'Raymax' x-ray unit. Four samples of clear quartz were studied: (i) the Brazilian quartz previously referred to, and (ii) three samples of unseeded crystals of synthetic quartz grown by the hydrothermal method (Hale 1948) at temperatures of 290° c., 380° c. and 390° c. In all cases sharp high-angle lines were obtained; the constancy of the parameters throughout the whole of the Brazilian crystal and one of the synthetic samples was established by studying several different specimens made from each. After averaging the Bragg angles in the manner previously described, the equations arising from seventeen lines with θ greater than 60° were weighted and a least squares calculation performed assuming the extrapolation to be linear in $\cos^2 \theta$.

In order to facilitate direct comparisons with earlier results, we have deliberately used Siegbahn's values for the x-ray wavelengths, i.e. $\text{Cu K}\alpha_1$, 1.537395 kx.; $\text{Cu K}\alpha_2$, 1.541232 kx. The parameters found at 25° c. were corrected to their values at 18° c., the reference temperature used by previous authors, and the coefficients of linear thermal expansion given by Jay (1933) were employed for this purpose.

Lattice-Parameters (in kx.)

Sample	25° c.			18° c.		
	<i>a</i>	<i>c</i>	<i>c/a</i>	<i>a</i>	<i>c</i>	<i>c/a</i>
Brazilian quartz	4.90309 ₃	5.39367 ₃	1.100057	4.90258 ₅	5.39333 ₃	1.10010
Synthetic quartz (290° c.)	4.90372 ₈	5.39415 ₈	1.100012	4.90321 ₉	5.39381 ₈	1.100056
Synthetic quartz (380° c.)	4.90334 ₇	5.39387 ₆	1.100038	4.90283 ₉	5.39353 ₆	1.100083
Synthetic quartz (390° c.)	4.90324 ₃	5.39384 ₄	1.100056	4.90273 ₅	5.39350 ₆	1.10010
Bradley and Jay (1933)				4.90288	5.39327	1.10002
Same (modified by Cohen)				4.90287	5.39344	1.100057
Lipson and Wilson (1941)				4.90320	5.39371	1.10004

The second set of values given under the names of Bradley and Jay are those determined from their experimental Bragg angles by Cohen (1935) by means of a least squares calculation. Miller and DuMond (1940) also measured the parameters of a sample of natural quartz and obtained values considerably below ours and, therefore, diverging even further from Lipson and Wilson's values. They used a focusing camera which had been calibrated by a direct method. When expressed in kx. at 18° c. their parameters are $a=4.90220_9$; $c=5.39330_6$; $c/a=1.10018$.

§ 4. DISCUSSION OF THE EXPERIMENTAL RESULTS

It is immediately obvious from the results that the parameters of quartz are not sufficiently constant to found a precise standard upon them. The overall variations in parameters so far measured with 19 cm. cameras is seen to be 0.011%, while the experimental error is at most 0.0015% in the results of the different observers. Bradley and Jay (1933) demonstrated that the impurities in smoky, rose and milky quartz crystals caused relatively large increments in the parameters. Their experiments on different clear samples could not, however, be regarded as final; they assumed that the same axial ratio, for which their value is now known to be too small, held for all the specimens and employed rather inadequate graphical extrapolations throughout. The weakness of such a procedure has already been indicated by the calculations of Cohen. The variations found may possibly be interpreted as being due either to impurities or to holes built into the lattice during growth, but nothing is known of the existence or characteristics of such effects in the crystals used.

It is rather remarkable that in the parameters at 18° C. which we have calculated the axial ratio varies quite markedly and decreases as the parameters increase. The elastic anisotropy of quartz suggests that if impurities are present and cause a small stress in the lattice which is approximately homogeneous, then such a behaviour of the axial ratio can be explained. To test this further the results for the four measured samples were taken in pairs and the fractional increment in each parameter for a given pair determined. The ratio of these increments was then compared with the known ratio of the strain components, $s_{11} = 1.269 \times 10^{-12}$ and $s_{33} = 0.971 \times 10^{-12} \text{ cm}^{-2} \text{ dyne}^{-1}$, parallel and normal to the optic axis (Cady 1946). The ratio of the increments normal to the optic axis to those along that axis should therefore be $1.269/0.971 = 1.307$ if the lattice is stressed isotropically. The experimental values were calculated and the six possible pairs of parameters gave 1.419, 1.438, 3.85, 1.675, 0.98, 1.453 for the ratio. The value 3.85 arises from a pair of parameters different from one another only by the order of the experimental error, and if it is neglected the mean ratio is 1.393. This agrees to within 10% with the theoretical value for homogeneous stress and, bearing in mind that these ratios are those of the increments in the fifth significant figure, this agreement is very striking. We may conclude that the results obtained may be interpreted as indicating that an impurity content in quartz produces a small stress in the lattice and consequent alterations in the parameters.

The apparent dependence of the parameters of the synthetic samples on growth temperature may be fortuitous; it is hoped to obtain further results later. The results for these crystals when plotted against the growth temperature lie on smooth curves.

ACKNOWLEDGMENTS

The author wishes to thank Dr. Danforth R. Hale, of the Brush Development Co., Cleveland, Ohio, for the samples of synthetic quartz used in the experiments. They were grown under contract from the U.S. Signal Corps. He is indebted to Dr. J. W. Mitchell and Dr. H. F. Kay for encouragement throughout the work, and also to the former for his assistance in the preparation of this paper. Finally, thanks are due to the Queen's University of Belfast and the Northern Ireland Ministry of Education for financial assistance.

REFERENCES

- BRADLEY, A. J., and JAY, A. H., 1932, *Proc. Phys. Soc.*, **44**, 563 ; 1933, *Ibid.*, **45**, 507.
 BRADLEY, A. J., LIPSON, H., and PETCH, N. J., 1941, *J. Sci. Instrum.*, **18**, 216.
 CADY, W. G., 1946, *Piezoelectricity* (New York and London: McGraw-Hill Book Co.), p. 137.
 COHEN, M. U., 1935, *Rev. Sci. Instrum.*, **6**, 68.
 EKSTEIN, H., and SIEGEL, S., 1949, *Acta Crystallogr.*, **2**, 99.
 HALE, D. R., 1948, *Science*, **107**, 393.
 JAY, A. H., 1933, *Proc. Roy. Soc. A*, **142**, 237.
 LIPSON, H., and WILSON, A. J. C., 1941 a, *Proc. Phys. Soc.*, **53**, 245 ; 1941 b, *J. Sci. Instrum.*, **18**, 144.
 MILLER, P. H., and DUMOND, J. W., 1940, *Phys. Rev.*, **57**, 198.
 NELSON, J. B., and RILEY, D. P., 1945, *Proc. Phys. Soc.*, **57**, 160.
 RYMER, T. B., and BUTLER, C. C., 1945, *Phil. Mag.* [7], **36**, 515.
 TAYLOR, A., and SINCLAIR, H., 1945, *Proc. Phys. Soc.*, **57**, 108.

LETTERS TO THE EDITOR

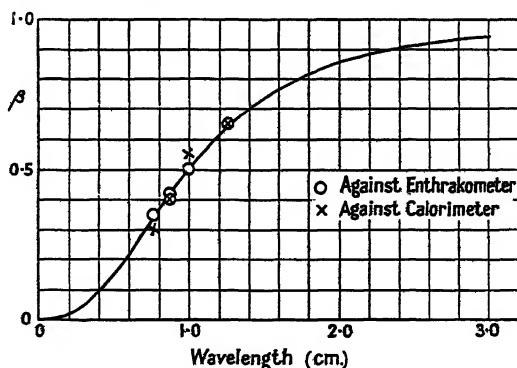
Discrepancies in the Measurement of Microwave Power at Wavelengths below 3 cm.

Recent experiments have shown that measurements of power at millimetre wavelengths, using a thermistor, may be in error by as much as 100% or more.

It has been the common practice at centimetre wavelengths to employ a water calorimeter for the measurement of high-level power (of the order of one watt or more). For low-level power, instruments have been used in which the power is absorbed in some form of resistance, the change in resistance being taken as a measure of the power. Instruments of this type are the bolometer, using a thin wire, the thermistor, using a semiconducting bead, and the enthrakometer (Collard 1946) using a resistive film.

At millimetre wavelengths the thermistor has been much employed, owing to its sensitivity and simplicity in use. Results obtained have been considered moderately reliable until recent experiments by one of us (J.C.) indicated a large discrepancy at a wavelength of 9 mm. between thermistor and enthrakometer measurements.

A water calorimeter, designed and constructed by G. Dalton of Telecommunications Research Establishment for the 7 to 12 mm. wavelength range and capable of measuring power levels as low as 20 mw. has made possible a direct comparison with the thermistor and enthrakometer. This comparison was carried out at T.R.E. over a wavelength range from $7\frac{1}{2}$ to $12\frac{1}{2}$ mm. at a power level of 20 mw. More accurate measurements were also made at 9 mm. of the enthrakometer against the thermistor for a power level of 1 mw. and of the enthrakometer against the calorimeter for 50 mw.; these measurements were consistent with the direct comparison of all three instruments. The source of power was provided by a range of reflection klystrons covering this wavelength band, made at T.R.E. and based on the original design of the Clarendon Laboratory, Oxford. The enthrakometer and water calorimeter agreed to within a few per cent, but the thermistor showed differences of the order of 100% as shown in the Figure.



Thermistor reading as fraction of actual power.

If the thermistor is considered as having a bead resistance R , shunted by a capacitance C , this combination being in series with the lead resistance r , it can be shown that the ratio β of the power dissipated in the bead to the total power is $\beta = 1/(1 + \omega^2 C^2 R r)$ provided $\omega^2 C^2 R^2 \gg 1$.

The effect of the capacitance C is to make the effective resistance of the bead very much smaller than R so that the lead resistance r becomes comparable to it.

Owing to the skin effect in the leads, the resistance r is proportional to \sqrt{f} . Assuming C and R independent of frequency we may therefore write $\beta = 1/(1 + kf^{2.5})$, where k is a constant for the bead.

An equation of this form has been fitted to the experimental points in the Figure. The agreement with the curve is good and requires $C\sqrt{(Rr)} = 5.3 \text{ pF}\Omega$ at a wavelength of 1 cm. An estimate of r (for 0.001 in. diameter 70% platinum—30% iridium wire, approximately 0.04 in. long) is about 2.5Ω at room temperature. The capacitance of a similar thermistor bead measured for us by the G.E.C. Research Laboratories, with a 45 Mc/s. bridge, was 0.2 pF. Assuming $R = 150\Omega$, which is the D.C. operating resistance of the bead, we obtain $C\sqrt{(Rr)} = 3.9 \text{ pF}\Omega$, which is the right order of magnitude.

From the above equation it is expected that β will decrease with increase in R . Altering the D.C. resistance of the bead over a range of 150 to 600 Ω (by controlling the bias current through the bead) produced a much smaller variation of β than would be expected if R were proportional to the D.C. resistance. This may indicate that the effective value for R for the radio-frequency power does not vary as rapidly with temperature as the D.C. resistance. If the thermistor bead is a polycrystalline material, the D.C. resistance will probably consist of the contact resistance of individual crystals. Hence at high frequencies these resistances will be shunted by the contact capacities and the specific resistance of the material will tend towards that of the individual crystals. In addition, minor effects may be produced by the variation of lead resistances and a possible variation of capacitance with bias current.

From the Figure it is evident that power measurements by a thermistor can be in error by a large factor at millimetre wavelengths. At 10 mm., for instance, the power measured by the thermistor is about one half the true value. Of course, a thermistor can still be used if it is calibrated and, since there are indications that thermistor beads of a particular type give fairly consistent measurements, it may be possible to dispense with individual calibration and use instead a simple correction factor.

Acknowledgment is made to the Chief Scientist, Ministry of Supply, the Director of Radio Research, Department of Scientific and Industrial Research and to the Controller of His Majesty's Stationery Office for permission to publish this note.

E.M.I. Research Laboratories Ltd.,
Hayes, Middlesex.

J. COLLARD.

Telecommunications Research Establishment,
Ministry of Supply.

G. R. NICOLL.
A. W. LINES.

24th January 1950.

COLLARD, J., 1946, *J. Instn. Elect. Engrs.*, **93**, Part III, A, No. 9.

REVIEWS OF BOOKS

Pulses and Transients in Communication Circuits, by COLIN CHERRY. Pp. xvi + 317.
1st Edition. (London: Chapman & Hall, 1949.) 32s.

This book, which is an introduction to the study of transients in networks, has been written with the object of bridging the gap between elementary A.C. theory and the methods of operational analysis, and in this the author has succeeded.

The scope is much wider than the title suggests, and there is little that cannot be profitably studied by anyone concerned with electrical communication, whether he is particularly interested in transients or not.

Among the topics discussed are the frequency spectra of various types of continuous and transient waves, Fourier series in real and complex form, conjugate vectors, Fourier integrals, equivalent and dual networks, the low-pass/band-pass analogy, the relations between the steady-state and transient response of a network, idealized networks, multi-stage amplifiers, echo effects in lines and networks, the method of paired echoes, and an introduction to transversal filters. In what is perhaps the best chapter of the book there is a thorough treatment of asymmetric sideband channels.

The author's method is "to use wherever possible rigorous physical arguments and only elementary mathematics. Mathematical language and notation are avoided as much as possible".

In fact, however, elementary mathematics and notation are freely used, and there are few pages that do not have a formula or so. The reader may sometimes wonder how rigorous the arguments are, but happily the results are usually correct.

It is clear that the book is full of important material, and it is therefore most unfortunate that this otherwise excellent work should be marred by a large number of errors. Some of these are obvious, but others require some research on the reader's part.

The definitions of the impedance and transfer characteristics on p. 85 are very strange; for example, the output impedance of a network is apparently equal to any impedance which the reader cares to connect across the output terminals.

Section 36—"Example of Fourier integral analysis of circuit response"—should be disregarded; only the final answer is correct.

The author does not appear to have made up his mind whether the "infinitesimal pulse" which he uses is a pulse of fixed amplitude and vanishingly small duration or whether it is of the Dirac type with a finite amplitude-time integral. The results obtained indicate the latter, but no clear definition is given and the two ideas are mixed in a bewildering fashion.

Section 47 deals with the suppression of the response which occurs in idealized filters before the signal is applied. The reader should ignore this Section. Both the method and results are wrong.

Distortionless and lossless lines are confused in Section 72. None of the formulae on pages 282-283 are valid for distortionless lines; some are true for lossless lines, others are simply not true.

These are only a few of the author's lapses. The list could be multiplied, but enough has been said to show that the reader, particularly if he is a beginner, will encounter several obstacles. It is difficult to avoid the conclusion that this book was written in a hurry, and it is to be hoped that the author will find time to make a thorough revision of both text and diagrams before a second edition is required.

A. S. GLADWIN.

Viscometry, by A. C. MERRINGTON. 1st Edition. Pp. viii + 142. (London: Edward Arnold & Co., 1949.) 16s.

This slim volume is intended for the practical research worker, being a concise yet up-to-date account of most of the methods available for measuring viscosity. The minimum of theoretical detail is given, but otherwise the layout closely follows that of Barr's *Monograph of Viscometry*. The generous bibliography given at the end of each chapter should prove adequate even for the specialist in any branch of the subject. The inclusion of apparatus dimensions in the appendix is welcomed, but it is hard to justify the inclusion of a table giving the viscosities of sugar solutions of various concentrations when, as the author points out, such a solution is not now recommended as a primary standard.

C. D.

Physics: Principles and Applications, by H. MARGENAU, W. W. WATSON and C. G. MONTGOMERY. Pp. x + 760. (New York: McGraw-Hill Book Co., 1949.) 30s.

According to the authors this book is written for American second year University students in engineering and the major sciences. In England, the book may be found useful to some entrance scholarship candidates and first year post-intermediate students. The book covers ground which is mainly not new, and perhaps an attempt to cover too wide a field has been made. At times very elementary subjects are discussed and then, rather suddenly, difficult topics are introduced. The accounts of well-known classical experiments are too brief if the reader is to get a real insight into the difficulties which had to be overcome.

For a first edition the book is remarkably free from printer's errors, but others of a more serious nature appear. Until these have been removed the book will tend to be misleading.

On p. 94 the expression $V \cos \theta$ is confusing, and somewhat earlier in the text the Greek word *δυναμὴς* is spelt *dynamos*. On pp. 159 and 326 a formula is given first with one sign and then with another. The modern tendency is to use a negative sign in expressing the

Newtonian law of gravitation. On p. 219 Poiseuille's formula is quoted and it is not pointed out that the formula given only applies to the flow of an incompressible fluid through a *horizontal* tube.

On p. 310 the term black-body radiation is preferred to full-radiation, and on the next page the formula $W = \epsilon \sigma (T_1^4 - T_2^4)$ is stated without any suggestion of a proof.

On p. 337 the authors use the term 'normal flux': surely flux is all that is required, for flux is $\int \vec{A} \cdot \hat{n} dS$, where the symbols have their usual meanings.

On p. 367 Ohm's law is given in a form which is not logical. The equation is true numerically and dimensionally, but will be most confusing. Later the importance of using a Weston standard cell at 0° C. is not indicated.

On p. 393 it is stated that Sir J. J. Thomson measured the specific charge of an electron in 1910: the first experiments were made in 1896 and the results published in 1897. On the following page it is stated that the same apparatus was used for the determination of the charge on a positive ion. This is the most serious error I have noted, for in the first experiments crossed-fields were used while in the second series they were parallel!

On p. 540 a diagram of a Kundt's tube shows (in the gas) a node at the end of the vibrating rod—a fault common to most textbooks.

The treatment of total internal reflection on p. 569 is inadequate and by no means satisfactory. An often-used conventional diagram is given, but the reviewer prefers one in which all the rays emanate from a point source, and, moreover, such diagrams should show that some light is reflected whatever the angle of incidence. The significance of the term "total internal reflection" then becomes more apparent. On p. 617 rays in a spectrometer are shown to have their origin at a point on the surface of a prism, and on the next page a ray suffering minimum deviation does not traverse the prism symmetrically. On p. 648 Newton's rings experiment is shown as Newton did it: it is preferable, however, to place the plate on top of the lens (cf. *Dictionary of Applied Physics*).

The book is well produced and the cost is not excessive, but a less pictorial cover might be more pleasing to many readers.

C. J. SMITH.

Applications physiques de la transformation de Laplace, by M. PARODI.

Pp. viii+177. (Centre National de la Recherche Scientifique, 1948.)

Price 17s. Obtainable from Messrs. H. K. Lewis & Co. Ltd.

In seven chapters, M. Parodi has written a valuable, and distinctly mathematical, survey of his subject. His first chapter is a treatment of the transformation in itself, and is notable for full mathematical rigour, but after this the author rather confuses matters by writing as though the symbolic calculus is identical with the use of the Laplace transformation. Actually, they are quite distinct, though the Laplace transformation provides one way of justifying certain formulae and practices in the other method.

The next three chapters deal respectively with the use of the method in solving ordinary and partial differential equations and in evaluating definite integrals. The fifth chapter applies the method to integral equations, and the sixth, the most advanced and most original, considers the relation between Volterra's *composition* and the Laplace integral. The last chapter is a treatment of electric networks, and steps somewhat outside the limits imposed by the title of the monograph, since it contains an introduction to the general subject of initial admittances and the like. It is thus the most closely related to physical problems.

The C.N.R.S. is doing a very fine work in publishing scientific material in a form somewhat better than we were used to in pre-war French printing. Here the type is clear, the paper good and the lay-out pleasing. The only misleading misprint noted is a y for g in equation (VI.1).

J. H. A.

Tables of Scattering Functions for Spherical Particles, by the U.S. NATIONAL APPLIED MATHEMATICS LABORATORIES. Pp. xiii+119. (Washington: U.S. Department of Commerce, National Bureau of Standards, 1949.) 45 cents.

These tables, prepared under the supervision of Dr. A. N. Lowan in the Computation Laboratory of the National Applied Mathematics Laboratories of the U.S. Bureau of Standards, are derived from the equations given in 1908 by Gustav Mie, which determine

rigorously, on classical electromagnetic theory, the diffraction of a plane wave by a sphere of any composition. They are addressed to physicists, chemists, biologists, and others who can use light scattering as a research tool: they originated from work on chemical warfare, and most of them were prepared on the recommendation of V. K. LaMer, professor of chemistry of Columbia University. Professor J. A. Stratton contributes an interesting historical introduction outlining the development of our knowledge of light scattering.

A fundamental parameter in the work is the ratio (α) $2\pi r/\lambda$ where r denotes the radius of the sphere and λ the wavelength of the incident light. Part I deals with the intensities of the two incoherent plane polarized components scattered by a transparent sphere illuminated with natural light, for α from 0.5 to 6.0; the ratio of the two intensities provides a measure of particle size from 0.05 to 0.2 micron; the angular distribution and total light scattered by a suspension of particles of uniform size can be used to determine their concentration. Angles of scattering at 10° intervals are considered and refractive indices (m) 1.33, 1.44, 1.55, and 2.00.

Part II gives K , the total scattering coefficient (that is, the total energy scattered per second per unit cross-sectional area of particle—illuminated at unit intensity) for $m=5$ and α from 0.5 to 12. Part III deals with the total scattering (and absorbing) coefficient K for absorbing materials (with extinction coefficient k between 0 and 0.1).

Part IV extends the original (LaMer) set of tables to certain complex values of m applicable to the scattering of microwave radar by suspended or falling water drops: these tables use dispersion measurements by J. A. Saxton of the National Physical Laboratory.

As Professor Stratton remarks in his introduction, the great labour expended in preparing this volume will yield ample benefits to physical and biological science, enabling quantitative discussion to replace qualitative ideas.

s.c.

Scientific Foundations of Vacuum Technique, by SAUL DUSHMAN. Pp. xi + 882. 1st Edition. (New York: John Wiley; London: Chapman & Hall, 1949.) 90s.

The introduction of Gaede's mercury diffusion pump marked the commencement of great advances in high-vacuum technique. A new industry—vacuum technology—has grown up, and a small-scale laboratory operation has been transformed within the past forty years into an industrial one, carried out on a scale that would have been almost inconceivable in the minds of the early workers in the field. Speed, and greater speed, is demanded for the exhaustion of all types of electronic equipment, cyclotrons, betatrons, etc., and for vacuum distillation, dehydration and the evaporation of metals and other substances. These higher speeds have been attained by paying more attention to jet design in pumps.

Only a few books have been written on the subject, and the first one to appear was that written by Dr. Dushman in 1922. The recent rapid advances have more than justified the present volume, which workers in this field have required for some time. In spite of its 882 pages, and of the fact that it contains a reference to the researches of practically every investigator in high-vacuum technique, it is no mere encyclopaedia, no catalogue of methods and results, but contains the theory which is so essential if successful operations are to be understood and appreciated. It is not only a complete revision of the earlier work, but it includes a very considerable amount of new material. The diagrams are numerous and good, and a useful feature is the copious references to original papers.

The opening chapters deal with the kinetic theory of gases and their flow through tubes and orifices, features which play important parts in high-vacuum work and which considerably affect pumping speeds. The treatment of mechanical and vapour pumps is interesting and explicit. The rôles of diffusion and condensation are discussed, and the extreme importance of jet design is fully explained. Alexander's modern type of pump is described and reasons are given why all vapour pumps do not function as diffusion ones—a conclusion which agrees with Langmuir's view on their operation. Manometers receive full treatment and the various indispensable leak detectors are included. Then follows a description of other methods whereby gases may be removed from vessels, such as the sorption of gases and vapours by solids, and in particular by activated charcoal, silicates, cellulose and metals. Diffusion through the latter and degassing problems are also fully

discussed. The chapter on chemical and electrical clean-up of gases at low pressures gives a comprehensive survey of the clean-up of gases by evaporated metals, technical getters, incandescent filaments, cold and hot cathode discharges, and the production of extremely low pressures in sealed-off devices. The later chapters deal with vapour pressures, rates of evaporation and dissociation pressures—including useful information on vacuum distillation—and the deposition of films.

The book is a complete treatment of the subject, one which should be in the hands of every worker in the fields of physics, engineering and chemistry.

F. H. NEWMAN.

An Introduction to Color, by R. M. EVANS. Pp. x+340. 1st Edition. (New York: J. Wiley & Sons; British Agents: Chapman & Hall, London, 1948.) 36s.

This is a book which should be read by every physicist who is interested in colour. So far as the purely physical aspects are concerned, he may find himself fairly familiar with those chapters dealing with light, light sources, illumination, spectral reflection and absorption, etc., although he will find these subjects dealt with in an instructive and lucid manner. He may be surprised at the complete absence of any algebra, yet he will discover most of the essential mathematics is present in the, to many people, more palatable form of graphs and diagrams. The outstanding virtue of the book, however, is in its breadth rather than its depth. Rarely, if ever before, has a book been written on colour in which the physics, physiology, psychology, aesthetics and technology of colour have been treated in such a balanced and sane manner. There is no suggestion that one approach is more correct or more important than another; on the contrary, the reader is left with a very clear impression of the inter-relation and inter-dependence of all these aspects. The physicist may even find it a humbling experience to read the book, because he can hardly fail to recognize that the physical aspect is the least subtle and least fascinating of any. He will, in particular, be introduced to many psychological factors which he had hitherto tended to ignore and to problems of perception of whose existence he was probably unaware.

Apart from its value to himself, the physicist will welcome this book as one which he can confidently recommend to his more artistically minded friends, and in particular to the art master at school. The physicist has a legitimate grievance that right from his earliest years the youngster at school learns about colour in a manner that is scientifically misleading. A boy discovers that when he mixes yellow and blue paints, he produces green, but he is quite unaware whether this result is due to the light or the pigment or the eye. From this initial half-conception comes much of the subsequent misunderstanding between the scientist on the one hand and the artist and craftsman on the other. We can recognize that it is not going to be easy to explain about spectral absorption to the boy or girl in his primary school, but we can at least expect the teacher himself to have some understanding of the problem. Evans' *Color* will give him most of what he needs to know in a manner which cannot fail to delight him.

This is undoubtedly a very fine book, refreshingly modern, beautifully produced and profusely illustrated. There are, no doubt, some details and omissions which might be criticized, but it would be churlish to dwell on these. We should, on the contrary, congratulate the author on a very fine achievement and hope for a wide and increasing circulation for his book.

W. D. WRIGHT.

Foundations of Modern Physics, by T. B. BROWN. Pp. xvi+391. 2nd Edition. (New York: John Wiley & Sons; British Agents: Chapman & Hall, 1949.) 30s.

Professor Brown's book is successful if its aim is to give to science students, in their last year at school or in their first year at the university, a sketch of the type of experiments and an outline of the ideas which form the basis of modern physics.

The production of the present work is excellent, the printing is good, the diagrams are clear and concentrate on fundamentals, and there are extremely few errors of any kind. There does not seem to be any way in which the treatment of so wide a subject

could be substantially improved within the limits of 400 pages. Yet it does not seem possible to deny that the analysis is shallow ; also the contribution of American workers is over-emphasized in the last few chapters. Perhaps the author can hardly be blamed for this since the advance of modern physics has been so rapid and on so wide a front that some selection is inevitable. Indeed, the reviewer feels that the days when a semi-popular exposition of physics can be undertaken are rapidly drawing to a close. No one expects advances in modern mathematics to be made generally intelligible or available to first-year students, and certainly the number and quality of those engaged in advancing modern physics is such that the breadth, content and complexity of the subject are likely to increase still further.

F. C. CHAMPION.

Corrigenda

“ Wave-front Aberrations of Oblique Pencils in a Symmetrical Optical System : Refraction and Transfer Formulae ”, by W. WEINSTEIN (*Proc. Phys. Soc. B*, 1949, **62**, 726).

Page 733, equation (14), the first term in the bracket on the right-hand side *should read*

$$\frac{3}{2} \sin I \cos I \left(\frac{\cos I}{r} - \frac{2 \cos^2 I}{t} \right) C_{30}$$

instead of

$$\frac{3}{2} \sin I \cos I \left(\frac{\cos I}{r} - \frac{\cos^2 I}{t} \right) C_{30}.$$

Page 737, equation (29), the left-hand side of the equation *should read*

$$\Delta(C_{32} N \cos^2 I) \quad \text{instead of} \quad \Delta(C_{12} N \cos^2 I).$$

On the right-hand side, the first term in the second line *should read*

$$+ \frac{3}{2r} C_{30} \sin I \cos^2 I \quad \text{instead of} \quad - \frac{3}{2r} C_{30} \sin I \cos^2 I,$$

and the first term in the third line *should read*

$$+ \frac{1}{4r^3} \cos I \quad \text{instead of} \quad + \frac{1}{8r^3} \cos I.$$

Reprints of Papers in the ‘Proceedings of the Physical Society’

In general, reprints of papers published in the *Proceedings of the Physical Society* are not available at the Offices of the Society, but may occasionally be obtained from the authors of the papers concerned. The address of the author is given at the head of each paper.

Exceptions to this are the special Lectures delivered before the Society, such as the Guthrie Lecture, the Presidential Address, the Charles Chree Lecture and the Rutherford Lecture.

CONTENTS FOR SECTION A

PAGE

Mr. E. B. ANDREWS and Dr. R. F. BARROW. The Band Spectrum of GeF	185
Mr. B. A. BILBY. On the Interactions of Dislocations and Solute Atoms	191
Dr. R. J. BENZIE and Dr. A. H. COOKE. Spin-Lattice Relaxation in some Paramagnetic Salts	201
Dr. R. J. BENZIE and Dr. A. H. COOKE. Specific Heats of some Paramagnetic Salts at Temperatures near 1° K.	213
Dr. D. G. E. MARTIN and Dr. H. O. W. RICHARDSON. The 'External' Photoelectron Spectrum and the Intensities of the γ -Rays of Thorium (B+C+C')	223
Dr. H. O. W. RICHARDSON. Ellis and Aston's Theory of the Height of a Photoelectron Line	234
Prof. L. JÁNOSSY. Note on the Fluctuation Problem of Cascades	241
Dr. NORA PAGE. Some Observations on the Nuclear Disintegrations caused by Cosmic Rays in Photographic Emulsions	250
Dr. K. J. LE COUTEUR. The Evaporation Theory of Nuclear Disintegrations	259
Mr. C. H. COLLIE, Dr. P. F. D. SHAW and Mr. H. J. GALE. The Correction of Specific β -Radioactivity for Self-Absorption	282
Dr. D. K. C. MACDONALD. The Magneto-Resistance of the Alkali Metals	290
Letters to the Editor :	
Mr. E. R. RAE. The Radiations of Praseodymium-142	292
Mr. R. M. LITTAUER. The Angular Distribution of Gamma-Rays from ${}^7\text{Li}(p, p){}^7\text{Li}^*$	294
Prof. K. SITTE. On the Production of Penetrating Showers in Nuclear Collisions	295
Mr. G. R. EVANS and Mr. T. C. GRIFFITH. Some Decay Effects of Mesons in Penetrating Showers observed in a High-Pressure Wilson Cloud Chamber	296
Mr. D. M. EVANS and Dr. H. WILMAN. Epitaxial Strain and Disorientation in Crystals Growing in Single-Crystal Substrates	298
Corrigendum	299
Reviews of Books	300
Contents for Section B	302
Abstracts for Section B	303

ABSTRACTS FOR SECTION A

The Band Spectrum of GeF, by E. B. ANDREWS and R. F. BARROW.

ABSTRACT. Two band-systems attributed to GeF have been observed in emission from a discharge through a stream of GeF₄ vapour. Vibrational analyses of both systems have been made and the following constants derived :

State	ν_e	ω_e	$x_e\omega_e$ (cm ⁻¹)
? ${}^2\Sigma$	35007.4	800.4	4.1 ₈
? ${}^2\Sigma$	23316.8	413.5	1.0 ₈
${}^2\Pi_{3/2}$	935	665.8	2.8 ₃
${}^2\Pi_{1/2}$	0	664.5	2.7 ₄

A brief discussion of the results, and a comparison between the force-constants of the mono- and tetra-halides of the group IV-B elements, is given.

On the Interactions of Dislocations and Solute Atoms, by B. A. BILBY.

ABSTRACT. An extension of the reciprocal theorem of Colonetti is given and with it a general expression derived for the elastic energy of a multiply-connected solid containing sources of internal stress and subjected to external forces. The terms in the energy which depend on the relative position of a selected source of stress with respect to the field of all the others can be separated. The expression is used to estimate approximately the interaction energy between parallel edge dislocations with arbitrary slip vectors, between parallel screw dislocations and between an edge dislocation and a solute atom causing a spherical distortion of the lattice.

Spin-Lattice Relaxation in some Paramagnetic Salts, by R. J. BENZIE and A. H. COOKE.

ABSTRACT. The magnetization of a paramagnetic salt is a relaxation process with a relaxation time depending on the external field and temperature. Measurements have been made with the object of determining relaxation times at various fields and temperatures in the liquid helium range. A description of the apparatus is given.

The results show that paramagnetic relaxation is a complex phenomenon, and suggest that in any one specimen it must be described in terms of a distribution of relaxation times rather than a single relaxation time. In general we find that relaxation times are of the order of 10^{-2} sec., with a temperature dependence according to a power of $1/T$ between 2 and 3, while increasing with the external field H according to a power of H between 0 and 1.

Specific Heats of some Paramagnetic Salts at Temperatures near 1° K., by R. J. BENZIE and A. H. COOKE.

ABSTRACT. The magnetic specific heats of gadolinium sulphate and of several salts of the iron group of transition elements have been measured at temperatures near 1° K. by a method depending on paramagnetic relaxation. The results obtained are analysed into contributions due to magnetic dipole coupling and interaction of the magnetic ions with the electric field of the crystalline lattice. Cases in which these contributions are insufficient to account for the observed specific heats are discussed.

The 'External' Photoelectron Spectrum and the Intensities of the γ -Rays of Thorium ($B+C+C'$), by D. G. E. MARTIN and H. O. W. RICHARDSON.

ABSTRACT. The spectra of the photoelectrons ejected from various targets by the γ -rays of $\text{Th}(B+C+C')$ have been measured and the emission probabilities of the γ -rays of energy 0.238, 0.277, 0.299, 0.510, 0.582, 0.726, and 0.859 Mev. have been estimated. Using the known intensities of the internal conversion lines, coefficients of internal conversion have been found. The four low-energy rays seem to be magnetic dipole in character with some electric quadrupole admixture, but the strong 0.582 Mev. ray of ^{208}Pb is predominantly electric quadrupole. The possible spins of the nuclear levels are discussed.

Ellis and Aston's Theory of the Height of a Photoelectron Line, by H. O. W. RICHARDSON.

ABSTRACT. The formula of Ellis and Aston relating the intensity of a γ -ray to the height of its photoelectron line produced from a thick target is revised using the quantum-mechanical theory of stopping power. The effects of multiple scattering, bremsstrahlung and anisotropic emission of the photoelectrons are discussed and a general comparison of thick and thin target methods is made.

Note on the Fluctuation Problem of Cascades, by L. JÁNOSSY.

ABSTRACT. Diffusion equations governing the fluctuations of nucleon cascades and also of electron-photon cascades are given. The electron cascade is treated at first in the usual approximation, but it is shown that the approximations are not essential for the treatment. The diffusion equations appear in the form of integro-differential equations which can be solved directly using numerical integration. It is hoped to present numerical solutions in a later publication.

Some Observations on the Nuclear Disintegrations caused by Cosmic Rays in Photographic Emulsions, by NORA PAGE.

ABSTRACT. An account is given of a study of 2,150 stars in Ilford C2 plates and 312 stars in Kodak NT4 plates exposed under different conditions. Possible values for the cross section for star-producing collisions are discussed. It is concluded that this collision cross section is equal to the geometrical cross section of the nuclei, and that the star-producing particles are non-catastrophically absorbed, leading to the observed absorption cross sections. It is shown that nuclear transparency could not account for the observed results. It is estimated that 17% of all stars and 40% of stars from which two or more lightly ionizing particles are emitted are caused by ionizing primary particles. The size distributions for stars caused by ionizing and non-ionizing particles are given, and possible explanations for the difference between them are considered. The experimental observations are in agreement with the theory that the star-producing particles are mainly neutrons and protons secondary to the primary cosmic rays.

The variation with star size of the proportion of doubly charged to singly charged particles is found, and for stars of more than six prongs is in agreement with Le Couteur's extension of the evaporation theory. A value of 0.54 is deduced for the proportion of doubly charged particles from disintegrations in the light elements of the emulsion.

An increase in the proportion of lightly ionizing tracks is observed for stars of more than eight prongs, and is ascribed to the creation of mesons.

The Evaporation Theory of Nuclear Disintegrations, by K. J. LE COUTEUR.

ABSTRACT. The relationship between nuclear temperature and excitation energy is discussed taking into account the thermal expansion of the nucleus and the effect of the neutron excess. Using these results, the evaporation theory is put into a form suitable for the treatment of highly excited nuclei. The competition between the emission of proton, neutron, deuteron, triton and α -particle is considered in detail, using Weisskopf's method of detailed balance.

Detailed numerical values are calculated for the disintegrations of silver bromide observed in photographic plates. The main points of comparison with experiment are (i) the energy distribution of the evaporated particles, and (ii) the ratio of emission of singly and doubly charged particles. The comparison yields close information about the relationship of nuclear energy and temperature; the agreement with theory is as good as can be expected. It is necessary to assume that the Coulomb potential barrier is reduced at high excitation energies.

The Correction of Specific β -Radioactivity for Self-Absorption by C. H. COLLIE, P. F. D. SHAW and H. J. GALE.

ABSTRACT. The self-absorption of β -rays in radioactive films between 5 and 100 mg/cm² superficial density has been investigated. The marked departures from the exponential absorption law are shown to be due to scattering of the β -rays in the film. The use of very thin films to avoid the correction for self-absorption is not recommended as a routine method.

The Magneto-Resistance of the Alkali Metals, by D. K. C. MACDONALD.

ABSTRACT. The magneto-resistive effect has been examined experimentally in the alkali metals at low temperatures. It is concluded that deviations from the free electron gas model become considerable in the case of the heavier metals, caesium and rubidium.

THE PROCEEDINGS OF THE PHYSICAL SOCIETY

Section B

VOL. 63, PART 4

1 April 1950

No. 364 B

Experimental Studies in Thermal Convection

By A. O. RANKINE *

33rd Guthrie Lecture, delivered 21st March 1949

Mr. President, Ladies and Gentlemen,

I feel that I ought to begin this lecture with a confession. When it was proposed last year that we should celebrate the seventy-fifth anniversary of our Society I discouraged the idea. This was partly because the number 75 did not seem to me round enough to make a fuss about, but even more because I knew from experience how much work the arrangement of the event might involve. My recollections of the Society's Jubilee in 1924 were still vivid. At that time I had recently been appointed Business Secretary, and thus had to shoulder much of the responsibility of organizing the proceedings, which lasted several days and included a series of elaborate functions, the most notable of which was the Jubilee Dinner at which our present King was the principal guest of honour. So now, out of sympathy for the present Officers, whose routine duties are so much more onerous than in my day, I wanted to save them from the additional labour of another celebration, which I judged to be premature.

I thought that my views would be shared by most of the members of the Society, but it appears quite definitely that I was wrong, and, anyhow, I have been overruled. For modesty forbids me from flattering myself that this large gathering is assembled here today merely to hear me discourse, and I know that several hundred Fellows and their friends intend to participate in the Soirée later on this evening. I suppose the truth is that the younger members of our Society particularly, for whom this celebration is the first, welcome the opportunities it affords for social intercourse which has been missed so much during the dark days of the war and its aftermath. And even those of us who are older must agree with them in believing that today's functions, limited in scope though they be for reasons of economy, will help us to get to know one another better, and foster our united efforts to make our Society flourish still more in the years that lie ahead.

Now, as a sort of penance for my former reluctance, it falls to my lot to open the proceedings by giving the thirty-third Guthrie Lecture on this day exactly seventy-five years after our Society was founded through Guthrie's influence on 21st March 1874. We have therefore a double reason for remembering him afresh and paying tribute to the energy and foresight he exercised in our interests during the formative years. We do so now sincerely in the presence of several

* This article will also appear in the May issue of *Proc. Phys. Soc. A*.

members of his family and other relatives, whom we welcome as on many former occasions, but with a note of deep regret that his daughter, Miss Hilda Guthrie, who has attended these lectures with great regularity and was present last year, has since died. We are especially pleased to be honoured again by the presence of another daughter, Miss Elfrida Guthrie, who has made a long journey to join us, in spite of her advancing age.

This would seem also to be an appropriate occasion for reviewing our Society's development from its foundation until now, with special emphasis on the last twenty-five years—indeed to make this the subject of my lecture. But this would be what the theologians would call a work of supererogation. For such a review has been done comparatively recently, and by one whose ability to tell the story effectively and elegantly is far superior to my own. I refer to our former President, Professor Andrade, whose presidential address in 1945 on "The History and Future of the Physical Society" I have found so refreshing to read again that I recommend all of you to do likewise.

The subject which I have chosen instead is one which, on account of its largely experimental character, would, I think, have had Guthrie's approval. Yet I have some doubts as to whether it will be worthy of its place in the Guthrie Lecture series. I have attended most of these lectures since they were instituted, and I recognize both the eminence of my predecessors and the scientific importance of the matters they dealt with. I have indeed a suspicion that I have been chosen as the lecturer this year more in recognition of my long official connection with the Society than for my contributions to physics. If that be so, I appreciate all the more the honour conferred on me, and the opportunity it may provide for putting on record in the *Proceedings* of our Society some of the results of the most arduous piece of work I ever did in my life, and in which I myself feel a certain pride in having performed. It was done during the war in collaboration with several colleagues, of whom I desire to mention particularly Mr. C. R. Young and Mr. C. Bird, in connection with the operation of dissipating fog over airfields. The results comprise comprehensive data relating to certain aspects of thermal convection in the air, and are at present accessible only in the documents of the now defunct Petroleum Warfare Department. Some at least of them I think may be found to be of more general interest and applicability, and it will be for the Editing Committee to say how much of the contents of my final report might properly be published by the Society.

It is not my intention to deal now with the question of fog clearance, except incidentally. My main purpose is to direct attention to some experimental facts that emerged concerning the behaviour of air heated artificially, and to the means whereby the facts were discovered. But to provide a suitable background for describing the work, and to explain how these experiments came to be undertaken, I must first recall the way in which the problem of dissipating fog by artificial means presented itself for solution as a war-time operation.

Towards the end of 1942 Mr. Churchill charged the Petroleum Warfare Department with the duty of providing means of dissipating fog over airfields. The method chosen was to heat the air by burning petrol close to the ground. Such a scheme had already been contemplated before the war, and experiments had been carried out by the Royal Aircraft Establishment under the direction of the Meteorological Committee of the Air Ministry. The results then obtained provided useful guidance, and there were available also the important calculations

made by Professor Brunt indicating the minimum fuel requirements, having regard to the fact that fog drifting across an airfield needs to be dissipated continually as it encroaches thereon.

The first step taken was to construct in the large unfinished reservoir at Staines a petrol-burning installation approximately on the required airfield scale. A similar installation with coke as the fuel was also laid out. Both these were used when atmospheric conditions were suitable to obtain data regarding the distribution of convective heat in the air, employing ordinary meteorological thermometers and anemometers for the purpose. But useful information was slow in accumulating, and it was decided to institute model experiments in which conditions would be under control. I was put in charge of this work and submitted my final report early in 1945.

What follows is mainly extracted from this report. As before indicated I have selected what appears to me likely to be of some general interest, not particularly in relation to the original problem. It being possible, however, that a few readers of this lecture may desire access to the whole report, I have deposited several copies of it in the Society's library where they will be available for perusal.

EXPERIMENTS IN THE EMPRESS HALL, EARL'S COURT, LONDON,
ON THE DISTRIBUTION OF HEAT FROM LINE BURNERS, IN
RELATION TO THE PROBLEM OF FOG CLEARANCE ON AIRFIELDS

Introduction

The process of clearing fog consists in establishing in the required regions conditions which lead to the re-evaporation of the water droplets constituting the fog. If the method employed is the thermal one the practical requirement is the delivery of heat in sufficient quantity to all those parts of the atmosphere from which the fog is to be removed by evaporation. In respect of airfields *rapid* fog clearance is also requisite, and this has in practice to be brought about by providing in general much more heat than is necessary merely to vaporize the water, so that, by sufficient drying of the atmosphere as well, the fog droplets vaporize quickly enough. Given the constitution of a particular fog—its temperature, its water content and the speed with which it is moving—the heat requirements for its rapid dissipation are easy to calculate.

A question obviously of fundamental importance is the manner in which the heat provided becomes distributed in the atmosphere. It has been held necessary to adhere strictly to certain conditions laid down by the air authorities, namely, that the sources of heat used—in practice lines of burning fuel—must be very close to the ground, and not nearer than 50 yards from the edge of the runway over which it is desired to clear fog. A further desideratum is that clearance should extend to a height of at least 100 feet above the centre of the runway. The problem thus became one of investigating the behaviour of the thermal convective jets originating in lines of burning fuel, under the influence of winds of various strengths and directions, including, of course, the special case of no natural wind at all.

In theory this problem could be solved by carrying out a large number of experiments in the open air on the actual scale of the proposed airfield installation. In practice, however, this would be a very lengthy and difficult process, as no control could be had over the atmospheric conditions, and observations would be

restricted on the occasions when those conditions happened to be favourable. It appeared, therefore, that experiments might with advantage be carried out on a much diminished linear scale, under artificial conditions fully under control and capable of reproduction whenever required. There were, in fact, good reasons for believing that, with proper adjustments of the rate of heat emission from the burning line, and of the speed of the imposed wind, a pattern of heat flow, resembling the full-scale pattern, could be produced and investigated with a model installation, the linear dimensions of which were, say, one-thirtieth or one-sixtieth of that used in the field. The sequel will show that this belief has been fully justified.

A further word of explanation is necessary before proceeding to describe the model experiments. Strictly speaking, the measurements of heat flow, whether in the model or in the field, ought to be made under atmospheric conditions involving actually precipitated fog. But the results obtained are nearly enough correct in a clear atmosphere, and, if necessary, an appropriate allowance can be made. Consequently, in the model experiments the observations throughout have been made without the inconvenience of having to create and maintain artificial fogs.

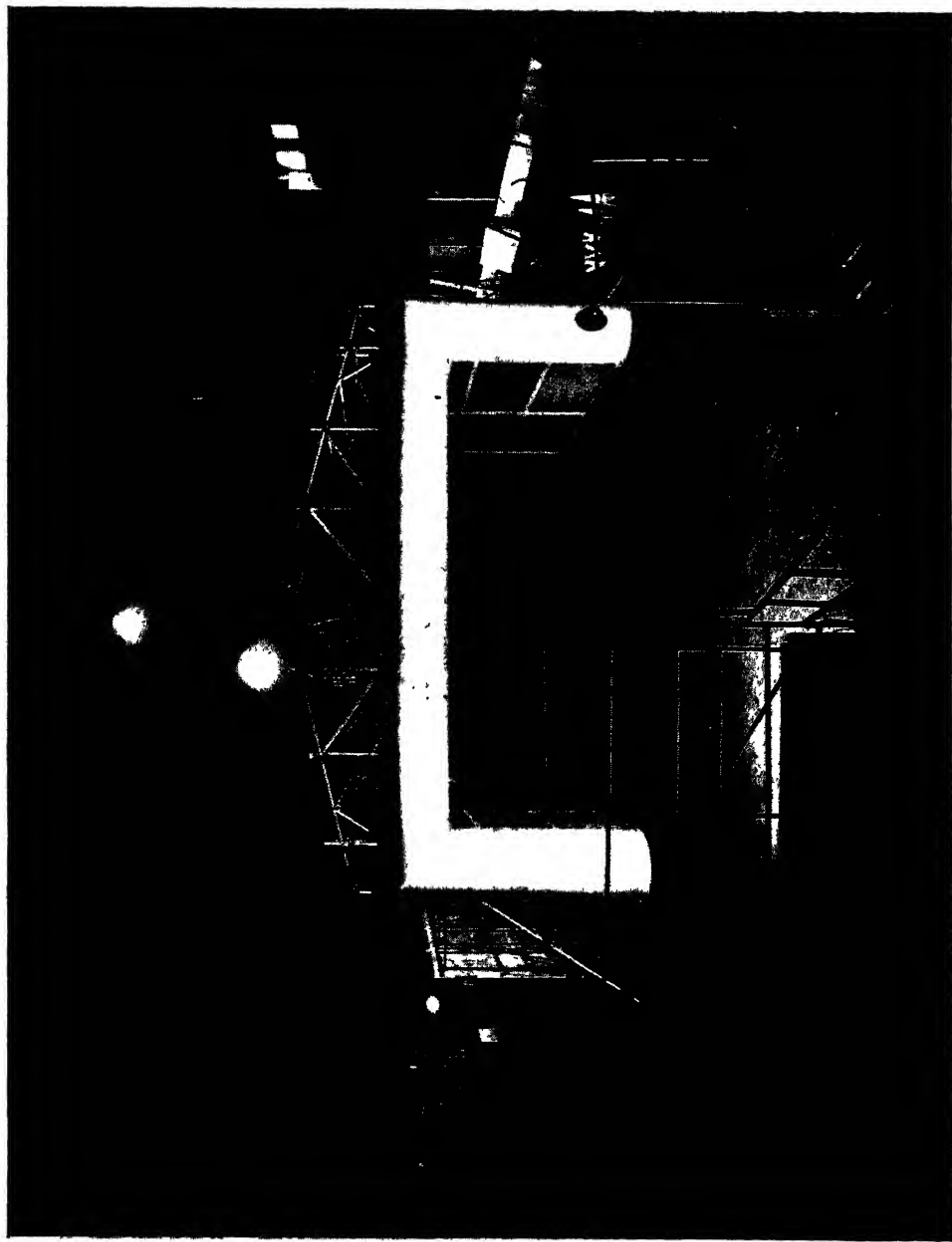
Wind Tunnel for Model Experiments

The model work has been carried out in the Empress Hall during the period from April 1943 to December 1944. For the purpose of most of the experiments a wind tunnel was constructed within the hall, which is a very spacious enclosure about 200 ft. long, 150 ft. broad and 70 ft. high. The tunnel itself was constructed with the advice and collaboration of the National Physical Laboratory, especially of Mr. L. F. G. Simmons, of that laboratory. It was built on the concrete floor of what had been a skating rink, and the main part was 100 ft. long, 30 ft. wide and 12 ft. high, the sides and roof being composed of wooden boards. At one end the rectangular cross section was flared down in a distance of 30 ft. to a square cross section of 12 ft. side, and thereafter to a circular channel 10 ft. in diameter which housed a 27 H.P. electric fan lent by the London Passenger Transport Board. At 30 feet inside the tunnel from the open end a wire mesh partition (30 meshes to the inch) was inserted for the purpose of rendering regular the air flow created by the suction of the fan. Doors in the sides gave access to the tunnel between the wire screen and the fan. Photographs of the exterior and interior of the tunnel are shown in Figures 1 and 2 (see Plates).

Investigation of Wind Structure in Tunnel

The range of air speeds required was from zero to a few feet per second, so as to correspond, with due allowance for scaling, with the winds likely to be operative in fogs on airfields. Running at its normal speed the fan made the air in the tunnel move much too fast. It was necessary, therefore, to reduce this by appropriate resistance adjustments on the fan motor. Further reduction was secured, as necessary, by shutting a pair of perforated doors housed in the square part of the tunnel just in front of the fan. In this way a range of air-speeds from 0.5 to 5 feet per second could be produced and reproduced as required. On a 1/60 linear scale, on which a large part of the work was done, this range corresponds to full scale values from 2.65 to 26.5 miles per hour.

Preliminary tests with a meteorological cup anemometer showed that the general air movement was fairly uniform in the tunnel except near the floor, roof and walls. Such anemometers were, however, not accurate enough for



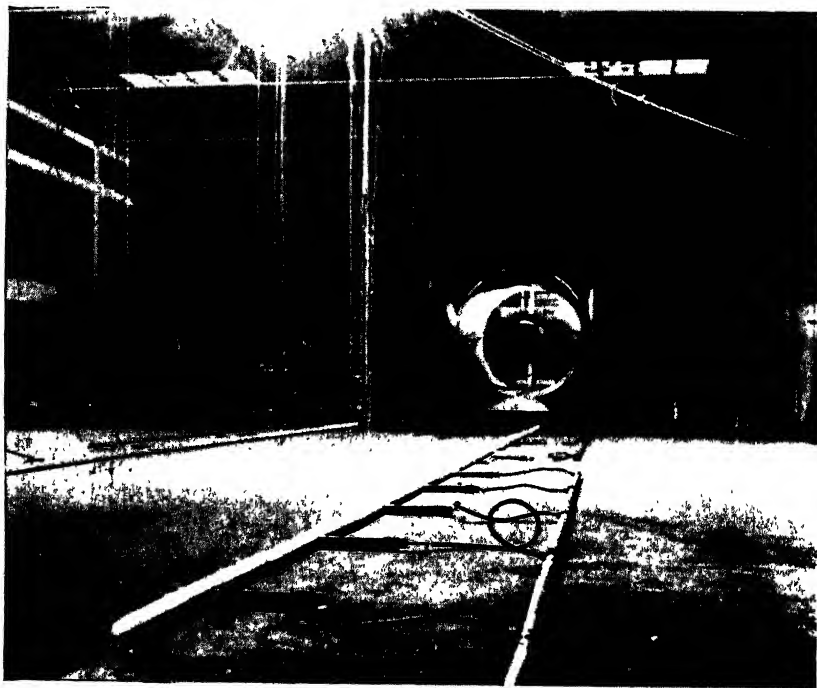


Figure 2. Interior view of wind tunnel showing fan and one arrangement of apparatus.

the purpose of measuring low wind speeds. Moreover, on the scale of the work they were too large, thereby interfering unduly with the air flow they would purport to measure. Recourse was had again to Mr. Simmons, who produced and made available about a dozen small electrical anemometers of a form which would not disturb appreciably the air currents under investigation. These anemometers will be described fully elsewhere. Here it must suffice to indicate that they depend on the principle that a thin wire electrically heated is cooled by air flowing past it to an extent which increases with the air speed, and that the cooling is measured electrically by thermocouple elements held close to the hot wires. These anemometers, besides being small and of suitable sensitivity, had the additional advantage of providing remote observation, so that the measuring operations could be performed outside the tunnel by means of instruments connected by thin wires to the anemometers disposed at suitable points inside.

Before being employed to investigate the wind structure in the tunnel under the various conditions of the experiments the anemometers had to be calibrated, so that the corresponding air speed could be derived from the measured electric potential of the thermocouple. In their original form the anemometers were calibrated at the National Physical Laboratory, but evidence soon accumulated that they changed sometimes in behaviour, so as to necessitate recalibration rather frequently. In order to do this without having to disconnect the anemometers and remove them from the tunnel—a somewhat laborious procedure also liable in itself to produce changes in them—a method was devised to re-calibrate them *in situ*. It consisted of establishing the air speeds along the centre of the tunnel, for various settings of the electric controls of the fan motor, by observing the times of passage between two fixed planes of almost weightless flakes of suitable material carried along by the air stream. The material used was the solid fuel known by the trade name of 'Meta', which is provided in small sticks. On touching a stick with a hot rod such as a soldering iron the material volatilizes and condenses again almost immediately in a form like snow flakes, but relatively much lighter. These fall very slowly, and in a horizontal air stream moving even as slowly as 0.5 foot per second, travel nearly horizontally. By releasing them in the tunnel and determining the average times of passage of many individual flakes between fixed planes 10, 20 or 30 feet apart, it proved possible to obtain reliable values of the air speed in the parts of the tunnel unaffected by the proximity of the floor, roof and sides. This was done for a range of air speeds from about 0.5 to 5 feet per second, and provided a convenient means of checking periodically the performance of the anemometers. Eventually Mr. Simmons modified the instruments into a form which was much more stable and reliable, but it was not until the later stages of the experiments that confidence could be placed on the results. The patterns of the wind structure were, in fact, re-determined in most cases by means of the anemometers in their final reliable form, the accuracy of measurement attained being a few per cent.

Measurement of Air Temperatures

Besides air speeds it was necessary to know the degree of heating of the air in order to define the complete pattern of the heat flow originating in the line burners. For this purpose electrical thermocouples were used so as to have again the advantage of remote observation outside the tunnel. Moreover, this kind of thermometer, consisting as it does of thin wires, could be disposed

in the air stream without modifying appreciably its structure. The installation of these thermocouples was facilitated greatly in the early stages of the experiments by the cooperation of the General Electric Company, and in particular of Mr. Lait, who spent many days on the work. The thermocouples consisted of soldered junctions of 30 s.w.g. copper wire and 32 s.w.g. Eureka wire, and they were calibrated so that the temperatures of air passing them could be deduced from the electrical potentials measured on a Tinsley potentiometer. As in the case of the anemometers, suitable switchboards enabled measurements to be made in turn on the various thermocouples mounted on slender frames at the points of investigation in the tunnel. At first measurements of the actual temperatures (above $0^{\circ}\text{C}.$) were made in the heated air and at a point outside the influence of the burning lines, the temperature rise being deduced by subtraction. Later the more convenient method was adopted of disposing the two elements of the thermocouple, one in the heated air and the other where there was no heat flow, so that the temperature rise could be obtained from a single measurement of potential. The accuracy of temperature measurement attainable was about $0.1^{\circ}\text{C}.$ or $0.2^{\circ}\text{F}.$ Investigation showed that radiation from the burner lines did not affect the thermocouples appreciably, and that they in fact assumed quickly enough the actual temperature of the air surrounding them.

The Burning Lines, and Method of controlling Heat Emission

The fuel used for the burning lines was butane, provided in iron bottles supplied from Petroleum Warfare Research Station, Langhurst. This fuel is especially convenient for small scale work because, the normal boiling point being about $0^{\circ}\text{C}.$, it usually develops in the bottles a vapour pressure somewhat above atmospheric, so that it is delivered as gas to the burners. Except occasionally on very cold days (when artificial warming of the bottles had to be resorted to) sufficient gas could be delivered to the burners at convenient pressures to provide the quantities of heat required, which, for the purpose in hand, had to range from zero up to a total value of about 5 therms per hour. The design and construction of suitable burners presented something of a problem. It was desired to burn the gas so that combustion was as complete as possible, for it was known that convective or sensible heat, as distinct from thermal radiation, was in fact the means whereby fog could be surely dissipated. Consequently, so-called non-luminous burning, such as occurs normally in a bunsen burner using coal gas, was aimed at. But ordinary bunsen burner nozzles and air inlets were not suitable, on account of the much greater (six times) calorific value of butane as compared with coal gas. New burners had to be made to meet the case, and, in the first instance, two pipes each 30 ft. long were provided from P.W.R.S., Langhurst, which through numerous orifices in the top about $\frac{1}{2}$ in. apart gave lines of non-luminous flames reasonably uniform in size over the whole length, although the butane entered the pipe through a suitable nozzle, and past a variable air inlet at one end only. These burners were used extensively in the earlier part of the investigation, but their dimensions were too great for all purposes. When laid on the floor, for example, the flames originated about 4 in. high, which, on the basis of model work on 1/60 field scale, would correspond to 20 ft. high in the field, whereas in airfield installations they start a foot or even less from the ground. To avoid this dimensional discrepancy, which would have vitiated the application of scaling in the manner hereinafter

described, the tunnel floor was in effect raised to the level of the flame orifices by boards mounted on blocks.

Later on a large number of smaller burners were made at the P.W.D. Experimental Station at Staines, some 10 ft. long and some 4 ft. long and all of 1 in. diameter, served by suitable nozzles obtained from the Physics Department of the Imperial College of Science and Technology. These burners proved to have a more uniform and quite satisfactory flame distribution, and, being relatively short, gave greater flexibility in constructing the various patterns of installation which had to be investigated. Moreover, on account of their small diameter, adjustment of floor level was not considered to be necessary. Another useful feature, as will appear presently, was that they could be altered, by means of the adjustable air inlet, so that the butane burnt luminously.

The various nozzles belonging to the burners were calibrated in relation to pressure above atmospheric and passage of butane. The pressure was measured by U-tube gauges, usually containing mercury, but occasionally water, and the rate of flow of butane was observed by a gas meter. With the nozzles thus calibrated serving any particular layout of burners, and the butane pressure being held constant by a control valve, any desired rate of delivery of butane per yard of burning line could be maintained. With the butane ignited, the calorific value being known, the corresponding gross heat emission rate per unit length could be deduced. The range in fact covered in the experiments was from 0.03 to 0.48 therms/yard.hour.

With the multiple burners sometimes used it was, of course, necessary to provide a system of service pipes connected with the bottles of butane and provided with sufficient taps to deliver gas to the individual burners. This was constructed as required out of 1 in. piping, and the connections to the burner nozzles were made with flexible rubber tubing. There was, however, no appreciable loss of pressure in the service system, the burner nozzles being so relatively small as to ensure that practically the whole pressure drop occurred across their orifices. Thus, with equal nozzles in burners of equal length, uniform heat output could be secured at a single pressure in all parts of the system.

Observational Procedure in Wind and Temperature Measurements

The procedure at first visualized was to lay out the particular type of burner system to be investigated, to instal the electrical anemometers and thermometers at suitable points, and then to make a few observations simultaneously of the air speeds and temperatures at these points both without and with the burners in operation. In practice this procedure had to be much modified in several respects. It was soon found that, owing to the turbulence of the air movement, the velocity and the temperature at a fixed point varied, sometimes very greatly, from moment to moment, and that, in order to be able to assign to the point reliable average values, a much larger number of readings would be necessary. Instead of a few readings usually fifty were taken, and the time occupied in thus dealing with every anemometer and thermocouple became much longer than had been anticipated. In all, several hundred thousand observations were made, each involving the manipulation of the appropriate potentiometer; and on some days more than two thousand data were added to the records. All these had to be corrected as necessary, classified and analysed before the final results emerged.

For reasons already mentioned, namely, the relative unreliability of the anemometers in their original form, it was not possible to carry out consistently the idea of simultaneous measurement of temperatures and air speeds. In some of the experiments for this reason no velocity measurements were made. In others, where a knowledge of velocities was more essential, they were re-determined during the latter stages of the investigation, when the improved anemometers became available. For example, what would naturally have been done first—the examination of variation of flow of the unheated air in the tunnel near the floor and sides—had to be thus postponed. Some of these vertical profiles of horizontal velocity are shown in Figure 3.

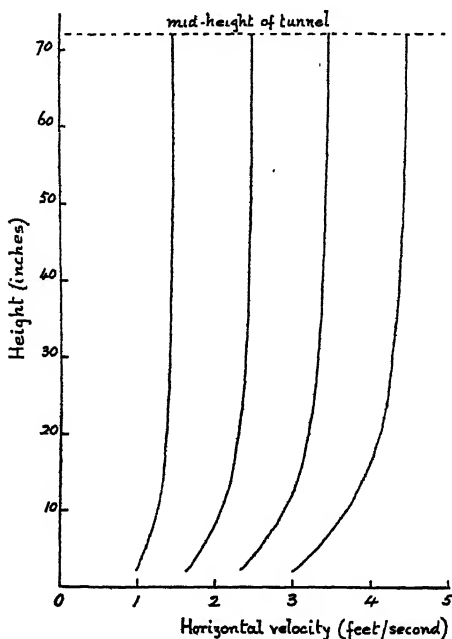


Figure 3. Vertical profiles of horizontal velocities of the air (unheated) in tunnel.

Detailed Investigation of Convective Heat Flow produced by Winds across Burner Lines

The principal feature of the work in the tunnel was the systematic investigation of the pattern of heat flow downwind of a burning line across which imposed winds of various strengths were operating. This particular case, in its full scale realization, is probably the most important in relation to fog clearance over runways, for it may be expected to provide answers to the questions—how much heat should be provided in a burning line, and where should the line be placed relative to the axis of the runway in order that fog clearance may be achieved to adequate heights over the middle of the runway, even when the wind originally blowing across both line and runway attains so high a value as, perhaps, 25 miles per hour?

Previous field experience has shown that in such circumstances the convective heat from a burning line is very easily deflected from the nearly vertical flow which occurs in air originally stagnant. Quite a low wind—a few miles per hour—

causes the heated jet to engage the ground on the downwind side of the line, while the upper edge makes an angle with the ground which diminishes rapidly as the imposed wind increases. Above this edge no appreciable amount of heat is imparted to the air, so that no fog dissipation can be expected there. Near the ground the required heat supply is more than adequate for the purpose; it is higher up that deficiency is liable to occur, and the aim is to ensure that at the minimum specified height for fog clearance the heat supply is at least enough. In other words the heat flow pattern under the influence of naturally occurring winds is not ideal. Such winds help in transporting the heat in the right direction from the burning line, but they give rise to a distribution of the heat which compels considerable waste at low levels in order to have at least enough at, say, 100 feet high over the middle of the runway.

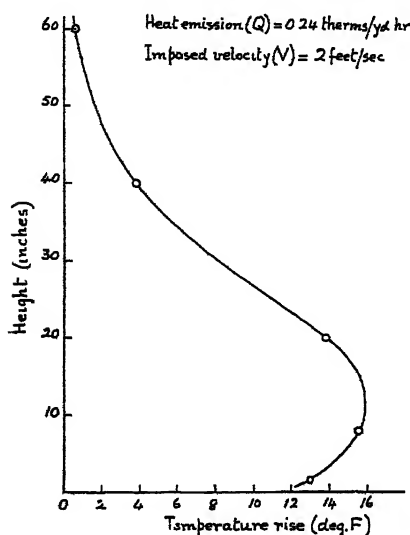
This was the problem now tackled with the model. A burner line was laid across the full width of the tunnel, usually about midway between the wire screen and the plane where the cross section began to flare down to the fan. Anemometers and thermo-elements were installed at suitable heights above the floor in the axial vertical plane of the tunnel at distances of 3 ft. 9 in., 7 ft. 6 in. and 15 ft. downwind of the burner line, and, as required, in the unheated upwind air as well. Systematic measurements of the temperatures and velocities of the moving air were carried out for a range of heat emissions from 0.03 to 0.48 therms/yard.hour, and of imposed winds from 1 to 5 feet per second at the mid-height (6 ft.) of the tunnel. Auxiliary measurements defined the vertical profiles of the various imposed winds when the air was unheated, as indicated in Figure 3, so that the changes caused by the burning operations could be determined. Also, traverses with anemometers and thermo-elements across the tunnel showed no appreciable variations except near the walls. Thus the results given for the axial vertical plane apply also to any parallel plane not too near the ends of the burning line.

These results have been analysed, interpolated and smoothed to remove evident errors in the customary manner, and are given in tabular form in the Appendix, with intervals small enough for it to be permissible to interpolate by proportional parts. A graphical method of representation which, *prima facie*, might have been deemed preferable, has been found too elaborate, on account of the multiplicity of the variations, since it would involve drawing some 380 graphs.

Discussion of Results for Winds across Burner Lines

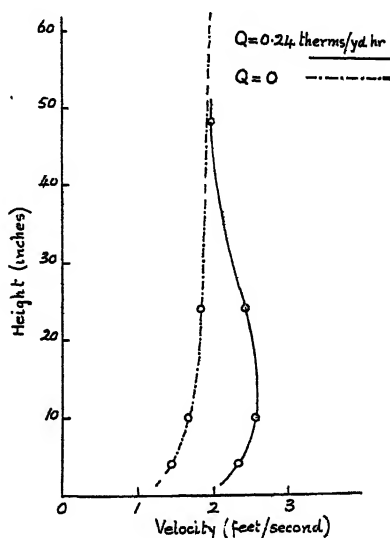
It is, however, instructive to represent a few of the cases graphically, and the chief points are thus illustrated in Figures 4, 5, 6, 7 and 8. Figure 4(a) shows the vertical profile of temperature rise of the air 7 ft. 6 in. downwind of the burning line when the gross heat emission is 0.24 therms/yard.hour, under the influence of an imposed wind of the magnitude 2 feet per second at the mid-height (72 in.) of the tunnel. This value implies a vertical profile of velocity of the type shown in Figure 3 for unheated air, and the particular profile in the present case appears in the dotted line in Figure 4(b), which exhibits also, by the full line, the profile modified by the heating of the air. Figure 5 represents a similar case, except that the imposed wind here is higher, namely 3.5 feet per second, with corresponding different profiles both heated and unheated. It will be noticed by comparison of Figures 4(a) and 5(a) that increase of wind speed reduces the

temperature rise at the higher levels and increases it at the lower levels. Also, inspection of Figures 4(b) and 5(b) shows, under burning conditions, that the original air-speed is enhanced considerably except at the higher levels, where there is little change. Figure 4(b) in fact demonstrates that, for low speeds of the imposed wind, the enhancement due to heating may cause the speed at low levels to exceed that at points higher up. This phenomenon of wind enhancement has been found to be generally true, and to be emphasized both by a reduction of imposed wind speed and an increase in the degree of heating.



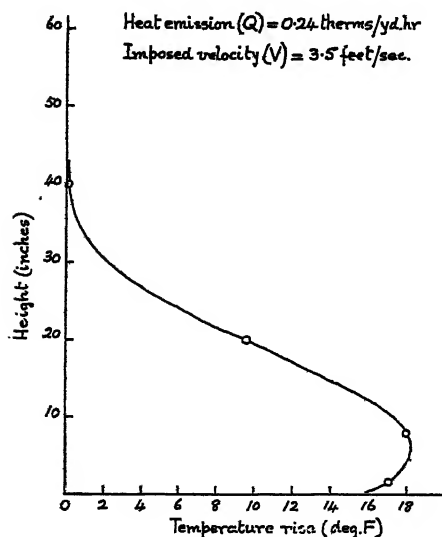
(a)

Figure 4(a). Temperature pattern 7' 6" downwind of burner.



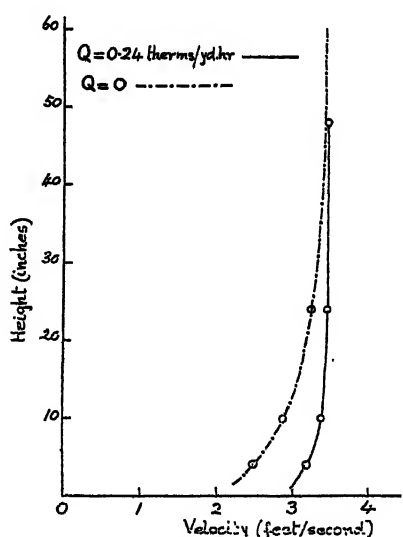
(b)

Figure 4(b). Velocity pattern 7' 6" downwind of burner.



(a)

Figure 5(a). Temperature pattern 7' 6" downwind of burner.



(b)

Figure 5(b). Velocity pattern 7' 6" downwind of burner.

The effect on the temperature rise of increasing the heat emission from the burning line is shown in Figure 6. Under the same imposed wind the temperature rise becomes greater at all heights, although not in a proportional manner, the ratio increasing rapidly with height. The heat emission in fact exercises control over the shape of the temperature profile as well as its dimensions. Another set of curves (Figure 7) illustrates the way in which the temperature

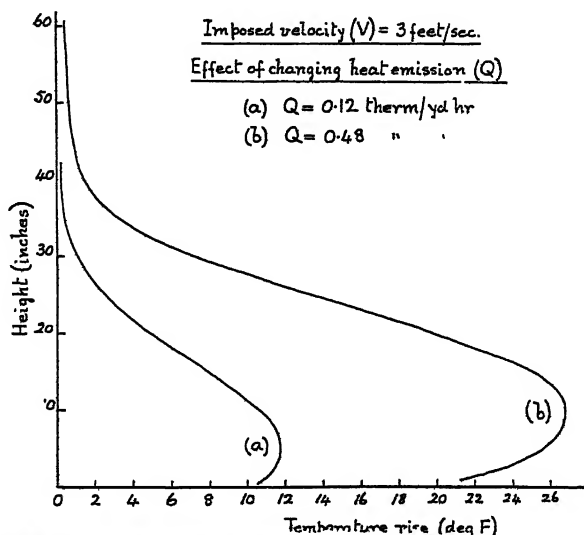


Figure 6. Temperature pattern 7'6" downwind of burner.

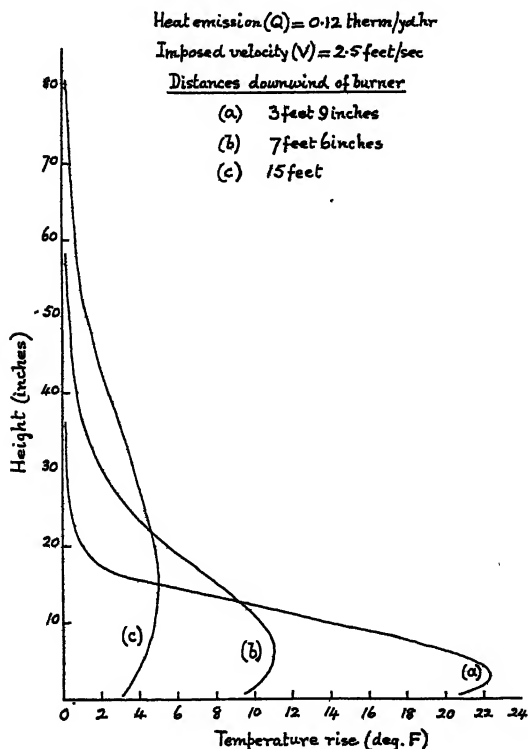


Figure 7. Temperature pattern downwind of burner.

profiles vary with distance downwind of the burning line, heat emission and imposed wind being constant. The most striking feature is the way in which the 'nose' of the curve (to which reference is made later) becomes blunted and elevated as the distance from the burning line increases.

Integration of Convective Heat Flow. Losses by Radiation and Ground Absorption

An important check can be made on the accuracy of the results by calculating the total quantity of convective heat passing downwind. This can be done when, as in most cases, the temperature rise is zero, or nearly so, at the maximum

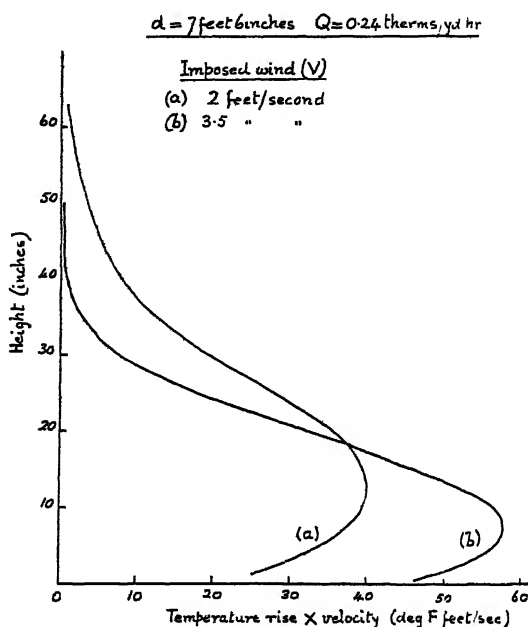


Figure 8. Heat flux downwind of burning line.

height of observation. The density and specific heat of air are known; the remaining factor involved in the determination is the product at all heights of the temperature rise and the wind speed, which is effectively horizontal. Figure 8 shows these products, plotted against height, for the data contained in Figures 4 and 5, and each of the two curves, for the same heat emission but different wind speeds, provide the means of graphical integration of the total convective heat flux, which is proportional to the area enclosed by the curve. The areas enclosed by the two curves are found to be nearly equal, implying that the convective heat flux is the same in the two cases. It is, however, considerably less than the gross thermal energy developed by combustion in the burning line. This is true, not merely for the two cases here considered, but quite generally. On the average only about 80 per cent of combustion heat has been found to cross any of the downwind planes of observation, in the form of convective heat, which alone could be effective in clearing fog.

This loss of about 20 per cent is accounted for in two ways. Some is lost to the floor, as is evident from the shapes of the temperature profiles. Always the temperature of the air close to the floor is lower than that somewhat higher up.

giving rise to the 'nose' of the profile. Moreover, the floor itself is always colder still, and this temperature gradient implies that the air does lose heat to the floor. The amount is not great; a fair estimate is about 5 per cent of the total thermal emission from the burning line, and most of this loss occurs closer to the burner line than 3 ft. 9 in., since there is no consistent evidence of variation of convective heat flux between this and greater distances downwind. It is probable that the ground of an airfield behaves in much the same way as the concrete floor of the tunnel with regard to transfer of heat from the air, so that the heat loss to the ground may also be taken as 5 per cent. It is fortunate that it is so small, and that most of the sensible heat remains in the air, where it is wanted for fog dissipation.

The remaining 15 per cent or so of the gross thermal energy of combustion, which does not appear as convective heat, is without doubt in the form of thermal radiation, including, of course, the small fraction of this which is luminous (even for so-called non-luminous flames). This is not appreciably absorbed by the air in the short distances (up to 15 ft.) over which measurements were made, and therefore does not become converted into sensible heat which would be registered by the thermometers. As regards full scale installations it may be reasonably assumed that the hydrocarbons constituting the petrol there used do not differ appreciably from the particular hydrocarbon butane in respect of radiation from flames. So that here also, unless the radiation is converted by absorption into heat capable of evaporating fog, it must be written off as a loss. The degree of this absorption is not known, but it may be said that conditions are somewhat more favourable in the full scale installation than in the model, owing mainly to the greater distances involved, and also to the fact that the constituents of a fog-laden atmosphere have a greater absorbing power than the clear unsaturated atmosphere in the tunnel, in respect of the specific radiation emitted by hydrocarbons burning non-luminously.

Radiation from Luminous and Non-luminous Flames—Experimental Comparison

This question of radiation was under constant discussion during the course of the model work at Empress Hall, and is the basis of the additional experiments now to be described. In the airfield installations the petrol has not been burnt non-luminously, but with less than the requisite supply of air for complete combustion, so that the flames have been highly luminous, and often near to being smoky. In this respect the burning lines have been different from the non-luminous (blue) flames of the butane ordinarily used in the model work. As mentioned earlier, however, it was possible, by restricting the air supply, to make the butane burn with luminous yellow flames more nearly similar to the petrol flames in the field. With this arrangement many measurements were made of the convective heat flux, for comparison with the corresponding flux for non-luminous burning. The results were of considerable significance. Generally, with equal rates of supply of butane to the burning line, the convective heat derived from luminous burning was found to be only about 80 per cent of that derived from non-luminous burning, i.e. 64 per cent of the gross amount calculated from the calorific value of the fuel. It appeared, therefore, that the loss by radiation was increased by a further 16 per cent if luminous burning was employed, making a total of about twice that for non-luminous burning.

This result was confirmed subsequently by direct comparisons made between the amount of radiation emitted by luminous and non-luminous flames consuming

equal quantities of butane, the same fraction of the total radiation from a line burner being received on a thermopile suitably disposed. With a butane consumption rate per yard equal to that in the former measurements of heat flux, the ratio of the total radiation from the flames when luminous to that for non-luminous combustion was found in a long series of observations to be 1.8, compared with the ratio 2 estimated indirectly as indicated in the previous paragraph. It was observed also that this radiation ratio depended upon the size of the flames, increasing as the flames were made larger by raising the fuel consumption rate. At the highest rate employed, when the flames were still only about 3 inches long, the radiation ratio was 2.4. This would imply, allowing also for the 5 per cent loss of heat to the ground, that the available convective heat is rather less than 60 per cent of the gross thermal emission. Compare this with the 80 per cent that can be obtained with non-luminous burning.

As regards the implications in full-scale petrol burning installations much the same conditions are likely to apply, or they may be even worse. The flames are much larger, perhaps two feet long, with a tendency to be smoky, consequently the degree of radiation from them will be even larger than in the model. Moreover, the additional radiation associated with their luminosity, occurring as it does in the near infra-red spectral region, is absorbed very little by any of the constituents of foggy air. Its transformation in the regions required to the only form of heat capable of dissipating fog, namely, convective or sensible heat, cannot, therefore, be relied upon to a useful extent. In fact, it is most unlikely that, with the petrol burners usually employed, the useful heat for purposes of fog clearance exceeds 60 per cent of the total calorific value of the petrol; it is probably considerably less when the burners are on the point of smoking.

The implication is clear. The petrol should be burnt non-luminously, in which case somewhat more than 80 per cent of the gross heat supply will be available for fog dissipation. An economic saving of 25 per cent of the petrol consumption could be achieved in this way.

Comparison of Model with full-scale Installation. Application of Principles of Dimensional Scaling

The measurements embodied in the tables of the Appendix provided the means of carrying out the crucial test as to whether, by the proper application of the principles of scaling (or dynamical similarity), the model results could be used to predict the behaviour of full-size installations, and thereby to specify the field requirements for the successful dissipation of fog. For this test some full-scale observations were necessary, and certain data of this character were available from field work at Staines.

Without going into the reasons thereof, the principles of scaling believed to be applicable may be simply stated as follows. Denote by L_m any particular length in the model (such as the distance of an observation point from the burning line) and by L_f the corresponding full-scale length. Similarly, V_m denotes the air speed at any point in the model, and V_f that at the corresponding point in the full-size installation. If also θ_m and θ_f are in like manner the corresponding temperature rises of the air, then

$$\frac{V_m^2}{L_m \theta_m} = \frac{V_f^2}{L_f \theta_f}$$

for all points. It is to be noted that this relation approximates to the truth only when θ_m and θ_f are small compared with the absolute temperature of the air, and this limitation must therefore be observed in applying it. The special application of the equation to the present argument deals with the case when

$$\frac{V_m^2}{L_m} = \frac{V_f^2}{L_f} \quad \text{or} \quad \frac{V_f^2}{V_m^2} = \frac{L_f}{L_m},$$

in which case $\theta_m = \theta_f$ or the temperature rises at corresponding points in the full-size installation and its linearly scaled model are the same. To realize this, having chosen the linear scale ratio L_m/L_f to be r , say, it is necessary to arrange that $V_m^2/V_f^2 = r$, or $V_m/V_f = \sqrt{r}$. We have control over the air velocities in the model in two respects, namely, the speed of the artificially produced wind across the burner, and the convective velocity of the thermal jet originating in the burner itself. As regards the latter, theoretical considerations* have shown that the velocities produced in the air by convective action are proportional to the cube root of the heat per unit length, H , of a long line source, which issues in convective form. Thus, to make the convective emission tally as between the full-scale installation and the model, the condition

$$H_m/H_f = V_m^3/V_f^3 = r\sqrt{r}$$

must be fulfilled in making comparisons.

The question then is this. If we take a series of measurements on full scale at Staines when a natural wind was blowing across a burning line, and select by the above rules corresponding observation points, imposed wind speeds and convective heat flux in the tables of the Appendix, do we in fact find in the model temperature rises equal to those of similarly placed points at Staines? This has been answered affirmatively by several comparisons carried out in the manner indicated, using data from Staines which appeared to be self-consistent and reliable. Figure 9, for example, relates to a petrol burning at Staines in which observations of temperature rise were made at several heights 75 yards downwind opposite the middle of a burning line 600 yards long. The scope of the model measurements was great enough to permit predictions by scaling for three different values of r , namely 1/60, 1/30, and 1/15. It will be observed that these predictions are reasonably consistent among themselves, indicating that internal scaling, i.e. within the range covered by the model investigation, is applicable. The curve represents the mean predicted temperature profile, and the closeness of the Staines observations to it is striking evidence of the validity of the scaling principles employed, even up to full scale.

Another result of the same kind is shown in Figure 10, which relates to a coke burning at Staines under like conditions as regard the relative positions of the observation points and the burning line. Here only two predictions from the model are possible, but they are fairly consistent, and the mean profile is again close to the actual Staines observations. It will be noticed that in both cases the temperature rises are small, the maximum value being 13.5°F., which is less than 3 per cent of the absolute temperature, so that the condition for the constancy of $V^2/L\theta$ is not violated. The difference between the predicted and actual values of the temperature rises nowhere exceeds 0.8°F., and this is well within the accuracy of measurement, having regard to the errors to which both field and tunnel observations were liable.

* Sir Geoffrey Taylor drew attention in 1943 to the work of W. Schmidt published in November 1941 in the *Zeitschrift für angewandte Mathematik u. Mechanik*, and himself contributed valuable notes on the subject of thermal and forced jets.

On the basis of these and other similar satisfactory comparisons it seems reasonable to conclude that the principles of comparison described are of general validity even for linear ratios as great as 60:1. In particular, the data contained in the Appendix may be confidently used in this way to discover the nature of the convective heat flow from a field burning line for all the higher cross-wind speeds likely in fog-clearing practice to be encountered, and for all heat emissions within reasonable practical attainment. Conversely, the requirements as to the position and heat emission of the burning line can be specified so as to ensure fog clearance over a runway to any requisite height.

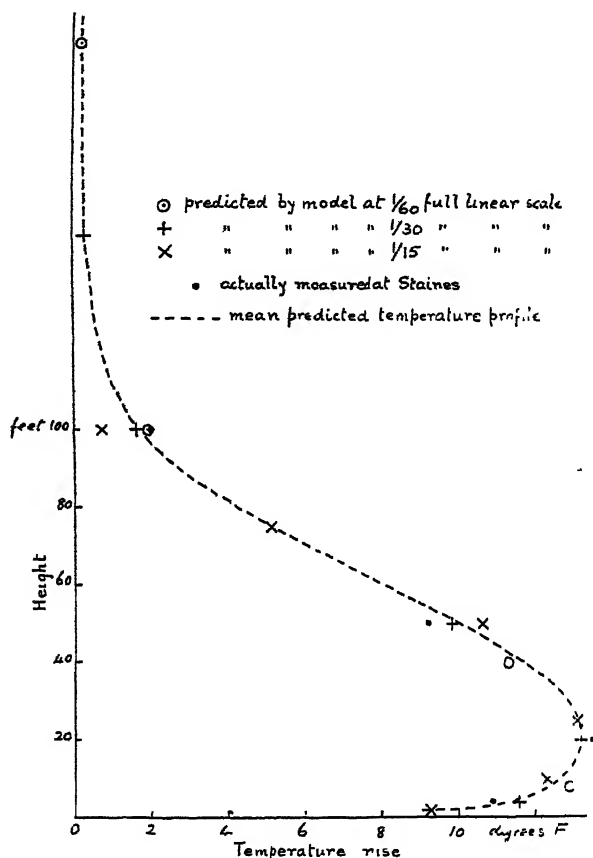


Figure 9.

Petrol burning, Staines, 14.12.42. Wind speed normal to burning line $8\frac{1}{2}$ m.p.h.; convective heat flux 22.4 therms/yd.hr.

Comparison of predictions from model observations with field results, using principles of dynamical similarity.

Limitation of Use of Tunnel in respect of Low Wind Speeds

Soon after the tunnel had been constructed early in 1943 Sir Geoffrey Taylor pointed out that its utility would be restricted, possibly seriously, by the existence of the roof, which would have no counterpart in the open atmosphere. He suggested modifications to transform the tunnel into what may be called a channel by removing the roof, and, by reversing the fan motion, to create the artificial wind by blowing instead of sucking. His proposals also contained other ancillary measures to render the velocity profiles suitable.

It turned out to be impracticable to implement these modifications, although, later, a separate channel was constructed and used, as described presently. The restrictions on the use of the tunnel had, therefore, to be examined and assessed.

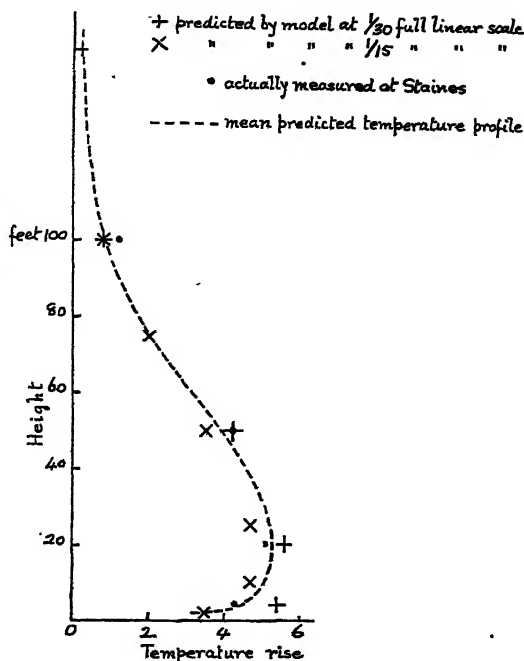


Figure 10.

Coke burning, Staines, 18.1.43. Wind speed normal to burning line 5 m.p.h.; convective heat flux 5.9 therms/yd. hr.
Comparison of predictions from model observations with field results, using principles of dynamical similarity.

Clearly, it would not permit the simulation of full-scale burning lines in calm atmospheres; for the heated air jet, rising vertically and reaching the roof, would be forced to spread out in *both* directions along the tunnel, in a manner which would not occur in the open atmosphere. The same would be true when the jet became deflected from the vertical by low-speed imposed winds; there would still be reversed air flow near the roof which would vitiate simulation of full-scale operations. The question to be decided was at what imposed wind speed this reversal, or re-circulation, ceased, owing to the thermal jet being depressed sufficiently not to engage the roof. This occurrence of re-circulation was tested both by observing visually the flight of meta flakes, and by observations with anemometers suitably disposed between roof and floor. If there is re-circulation the reversal near the roof is accompanied by increased velocity in the original direction near the floor, with a zero speed somewhere in between.

The result of these tests was to show that re-circulation was more liable to occur with the higher values of the heat emissions employed, as would be expected. Even with these heat ratings, however, it ceased to be noticeable for imposed wind speeds somewhat in excess of 1 foot per second. Consequently, the application of the tunnel experiments to predict open-air results has been limited in this sense, and the tables in the Appendix have been deliberately

restricted at the lower limits of V . To be on the safe side the lowest value of V included for the range up to $Q = 0.20$ therms/yard.hour is 1.5 feet per second, and for the range of Q from 0.24 to 0.48 therms/yard.hour to 2 feet per second.

This completes the description of the part of the work at Earl's Court which I deem to be suitable for publication in the *Proceedings*. For the rest, reference must be made to the full report, copies of which, as already stated, have been deposited in the Society's library. But I would like to indicate briefly here two aspects of this additional work which may find application in other directions in the future.

The first relates to the channel already mentioned as having been constructed with no roof in an attempt to realize conditions comparable with the open air. The difficulty here was to produce a reasonably uniform horizontal flow of air along the channel. Motor blowers directed along the channel proved to be abortive. Consideration was given to other possible methods, in particular that of using across one end of the channel a thermal line jet which, by its sideways entrainment of air, might be expected to create a horizontal wind of good uniformity. Theoretically, according to Schmidt, the entrainment speed would be independent of height, and, over a considerable distance along the channel, would vary little. This promising system could not be used because it required much more fuel than could be provided, but it remains one I should like to see implemented. Instead, a somewhat similar system was employed. Motor blowers of cold air directed upwards were installed at one end of the channel so as to constitute a line jet, thereby creating a horizontal drift of the air eventually entrained in this jet.

The other aspect of the work to which I wish finally to refer is the use of a device proposed and demonstrated early in 1944 by the General Electric Company's staff (Mr. Ryde in particular) to study visually the behaviour of thermal jets. This device, which has been called the 'Shadowgraph', consists of projecting through the jet a beam of light, originating in a small but very brilliant source, on to a white screen suitably situated beyond the jet. The result is that a pattern of fluctuating lights and shades appears on the screen, due to the point to point changes of the refractivity of the heated air. This pattern is very instructive, for, while not a complete substitute for the systematic investigation of the temperature and velocity structure of the jet, it does provide a quick means of getting a general picture of the movements of the heated air associated with various arrangements of burning lines. Through the courtesy of the General Electric Company the shadowgraph was used extensively during the later stages of the investigation. Had it been available earlier the planning of the work would have been greatly facilitated.

It proved to be possible to make some motion picture films of the patterns appearing on the shadowgraph screen in the tunnel. These, although by no means so brilliant as the originals, are reasonably good. By the nature of things, they cannot be included in this publication, and I am therefore glad to have been able to exhibit them, during the course of this lecture, through the courtesy of the Ministry of Supply.

REFERENCES

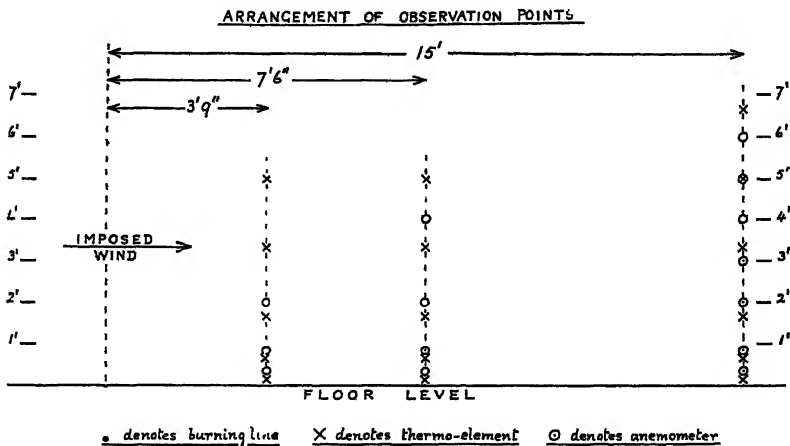
- ANDRADE, E. N. DA C., 1945, *Proc. Phys. Soc.*, **57**, xxi.
 BRUNT, D., 1936, *Trans. Faraday Soc.*, **32**, 1264; 1939, *J. Sci. Instrum.*, **16**, 137.
 SIMMONS, L. F. G. 1949, *J. Sci. Instrum.*, **26**, 407.

APPENDIX

EXPERIMENTS IN THE WIND TUNNEL AT EMPRESS HALL

Structure of convective air jets, originating in a burning line of non-luminous butane flames, under the influence of winds blowing normally across the line, as derived from observations of velocities and temperatures of the flowing air.

- V denotes the air speed at the mid-height of the tunnel (72 in.) when no heat is being injected. This defines the structure of the air current imposed on the thermal jet emerging from the burning line.
- Q denotes the thermal emission rate per unit length of the burning line, calculated from the calorific value of the butane.
- h denotes the height of the point of observation above the floor.
- d denotes the horizontal distance downwind of the burning line.
- θ denotes the excess of temperature of the air in the thermal jet at the point indicated over that of the air upwind of the burning line.
- v denotes the local horizontal air speed. The values of v in relation to those of h for $Q=0$ show the initial vertical velocity profile of the unheated air current. The heat-modified velocity profiles are derived similarly for the corresponding Q values.



The data given are derived from the averages of multiple observations, of which there were at least 50 for each experimental point. The temperature rises recorded are believed to be accurate to within 0.5°F. , or 5 per cent, whichever is the greater. The errors of the velocities shown are estimated not to exceed ± 5 per cent.

The chief purpose of the tabulated data of these model experiments is to provide means of predicting the conditions applicable to installations on other scales, by use of the accepted principles of dynamical similarity, namely, the constancy of the dimensional ratio $[V^2/L\theta]$, when θ is small, with the further relation that the fraction H of Q which is in the form of convective heat is proportional to V^3 . For these tables $H=0.8Q$, for the reasons given in the text relating to losses by radiation and ground absorption.

An apology is perhaps due for the units employed; the c.g.s. system would probably be preferred in some aspects of possible future use of the tables. The choice was determined by the fact that the data were required primarily for comparison with results obtained in the field, where British units had already been adopted for the full-scale work.

Tables of θ (in $^{\circ}\text{F.}$) for various values of d
(Q in therms/yard.hour)

$d=3\text{ ft. }9\text{ in.}$		$h=60\text{ inches}$							
Q		V in feet per second							
		1.5	2.0	2.5	3.0	3.5	4.0	4.5	5.0
0.00		0.0							
0.04		0.0							
0.08		0.1	0.0						
0.12		0.2	0.1	0.0					
0.16		0.4	0.2	0.0					
0.20		0.6	0.3	0.1	0.0				
0.24			0.5	0.2	0.1	0.0			
0.28			0.6	0.3	0.2	0.1			
0.32			0.7	0.4	0.3	0.2	0.0		
0.36			0.8	0.5	0.4	0.3	0.1	0.0	
0.40			1.0	0.7	0.5	0.4	0.2	0.1	0.0

		$h=40\text{ inches}$							
Q		V in feet per second							
		1.5	2.0	2.5	3.0	3.5	4.0	4.5	5.0
0.00		0.0	0.0						
0.04		0.1	0.0						
0.08		0.3	0.1	0.0					
0.12		0.5	0.2	0.1					
0.16		0.8	0.4	0.2	0.0				
0.20		1.1	0.6	0.3	0.1	0.0			
0.24			0.8	0.5	0.2	0.1	0.0	0.0	
0.28			1.0	0.7	0.4	0.2	0.1	0.1	0.0
0.32			1.3	0.9	0.6	0.4	0.3	0.2	0.1
0.36			1.6	1.2	0.8	0.6	0.5	0.4	0.3
0.40			2.0	1.5	1.1	0.9	0.7	0.6	0.5

		$h=20\text{ inches}$							
Q		V in feet per second							
		1.5	2.0	2.5	3.0	3.5	4.0	4.5	5.0
0.00		0.0	0.0	0.0	0.0				
0.04		1.1	0.8	0.4	0.0				
0.08		3.0	1.8	0.7	0.1				
0.12		4.9	2.9	1.1	0.2	0.0			
0.16		6.9	3.9	1.6	0.5	0.2	0.0	0.0	
0.20		8.9	5.0	2.2	1.0	0.5	0.2	0.1	0.0
0.24			6.1	3.0	1.5	0.8	0.4	0.3	0.2
0.28			7.3	3.9	2.0	1.1	0.6	0.5	0.3
0.32			8.7	5.0	2.6	1.3	0.9	0.7	0.5
0.36			10.4	6.2	3.1	1.5	1.1	0.8	0.6
0.40			13.0	7.7	3.6	1.8	1.3	0.9	0.7

Tables of θ (in $^{\circ}\text{F.}$) for various values of d (cont.) $(Q$ in therms/yard.hour) $d=3$ ft. 9 in. $h=8$ inches

Q	V in feet per second							
	1.5	2.0	2.5	3.0	3.5	4.0	4.5	5.0
0.00	0.0	0.0	0.0	0.0				
0.04	8.4	7.0	5.6	4.1				
0.08	14.1	13.1	11.6	9.6	7.5			
0.12	18.8	18.1	16.6	15.0	13.9			
0.16	23.0	21.8	20.3	18.4	16.9			
0.20	26.7	25.5	24.1	22.0	20.0			
0.24		29.4	28.0	25.7	23.2			
0.28		33.4	32.0	29.3	26.5	24.6	23.0	21.6
0.32		37.5	35.8	32.8	30.0	28.0	26.3	25.0
0.36		41.7	39.5	36.3	33.6	31.4	29.6	28.4
0.40		45.9	43.1	39.7	37.2	34.8	32.9	31.8

 $h=1.6$ inches

Q	V in feet per second							
	1.5	2.0	2.5	3.0	3.5	4.0	4.5	5.0
0.00	0.0	0.0	0.0	0.0				
0.04	9.8	10.2	10.6	11.0				
0.08	15.0	16.3	17.5	18.8	20.0			
0.12	19.5	20.8	21.9	22.9	23.7			
0.16	23.1	24.2	25.4	26.3	27.1			
0.20	26.3	27.6	28.9	29.8	30.6			
0.24		31.1	32.4	33.2	34.0			
0.28		34.7	36.0	36.7	37.5	38.3	39.1	39.9
0.32		38.1	39.6	40.3	41.1	41.9	42.7	43.6
0.36		41.3	43.1	43.8	44.6	45.4	46.3	47.3
0.40		44.5	46.3	47.3	48.1	49.0	49.9	50.6

 $d=7$ ft. 6 in. $h=60$ inches

Q	V in feet per second							
	1.5	2.0	2.5	3.0	3.5	4.0	4.5	5.0
0.00	0.0	0.0						
0.04	0.2	0.1						
0.08	0.4	0.2	0.0					
0.12	0.6	0.3	0.1	0.0				
0.16	1.0	0.4	0.2	0.1				
0.20	1.4	0.5	0.3	0.2	0.0			
0.24		0.6	0.4	0.2	0.1			
0.28		0.7	0.5	0.3	0.1			
0.32		0.8	0.6	0.3	0.1			
0.36		1.0	0.7	0.4	0.2	0.0		
0.40		1.1	0.8	0.4	0.2	0.1		
0.44		1.2	0.8	0.5	0.3	0.1	0.0	
0.48		1.4	0.9	0.5	0.3	0.2	0.1	0.0

Tables of θ (in $^{\circ}\text{F.}$) for various values of d (cont.)

(Q in therms/yard.hour)

$d=7$ ft. 6 in.

$h=40$ inches

Q	V in feet per second							
	1.5	2.0	2.5	3.0	3.5	4.0	4.5	5.0
0.00	0.0	0.0	0.0	0.0				
0.04	0.6	0.3	0.1	0.1				
0.08	1.4	0.7	0.4	0.2				
0.12	2.4	1.1	0.8	0.2				
0.16	3.6	1.7	1.0	0.3	0.0			
0.20	5.5	2.8	1.2	0.3	0.1			
0.24		3.8	1.5	0.4	0.1	0.0		
0.28		4.8	1.9	0.5	0.2	0.1	0.0	
0.32		5.7	2.3	0.7	0.4	0.2	0.1	0.0
0.36		6.6	2.8	0.9	0.6	0.3	0.2	0.1
0.40		7.6	3.3	1.1	0.8	0.5	0.4	0.3
0.44		8.6	3.8	1.4	1.0	0.7	0.6	0.5
0.48		9.6	4.3	1.8	1.2	0.9	0.8	0.7

$h=20$ inches

Q	V in feet per second							
	1.5	2.0	2.5	3.0	3.5	4.0	4.5	5.0
0.00	0.0	0.0	0.0	0.0				
0.04	3.5	2.6	1.7	0.8				
0.08	6.5	5.2	4.1	2.9	1.7			
0.12	8.4	7.9	6.5	5.0	3.6			
0.16	10.3	9.9	8.4	7.0	5.8			
0.20	12.0	11.9	10.2	9.0	7.8	6.3	4.9	3.5
0.24		13.8	12.0	10.6	9.6	8.1	6.6	5.0
0.28		15.5	13.9	12.5	11.3	10.0	8.4	6.5
0.32		17.5	15.5	14.2	12.8	11.5	9.9	8.0
0.36		19.0	17.3	15.9	14.3	12.9	11.1	9.4
0.40		20.7	18.9	17.4	15.6	14.1	12.5	10.9
0.44		22.3	20.5	18.8	17.0	15.4	13.7	12.2
0.48		23.9	22.0	20.0	18.1	16.5	14.9	13.5

$h=8$ inches

Q	V in feet per second							
	1.5	2.0	2.5	3.0	3.5	4.0	4.5	5.0
0.00	0.0	0.0	0.0	0.0				
0.04	5.1	4.9	4.6	4.2				
0.08	8.0	7.5	7.9	8.3	8.5			
0.12	10.4	9.8	10.9	11.4	11.5			
0.16	12.0	11.8	13.0	13.5	13.7			
0.20	13.4	13.6	15.0	15.5	16.0			
0.24		15.5	17.0	17.5	18.0			
0.28		17.4	18.7	19.4	20.0	20.0	19.8	19.6
0.32		19.0	20.4	21.0	21.5	21.4	21.2	21.0
0.36		20.5	21.9	22.5	23.0	23.2	23.3	23.1
0.40		22.0	23.1	24.0	24.5	24.7	24.8	24.6
0.44		23.4	24.5	25.3	25.7	25.9	26.0	25.8
0.48		24.6	25.7	26.5	27.0	27.2	27.2	26.9

Tables of θ (in $^{\circ}\text{F.}$) for various values of d (cont.) $(Q$ in therms/yard.hour) $d=7$ ft. 6 in. $h=1.6$ inches

Q	V in feet per second						
	1.5	2.0	2.5	3.0	3.5	4.0	5.0
0.00	0.0	0.0	0.0	0.0			
0.04	4.9	4.4	5.1	6.2			
0.08	7.0	6.7	8.4	9.1	10.8		
0.12	8.5	8.4	10.0	11.1	11.9		
0.16	9.9	10.0	11.6	13.0	13.7		
0.20	11.0	11.5	13.4	14.8	15.5		
0.24		13.0	14.9	16.4	17.1		
0.28		14.5	16.1	17.7	18.7	19.5	20.0
0.32		15.6	17.3	19.0	20.0	20.8	21.4
0.36		16.8	18.3	20.0	21.1	21.9	22.5
0.40		17.6	19.1	20.9	22.0	22.8	23.5
0.44		18.5	20.0	21.5	22.7	23.6	24.3
0.48		19.1	20.5	22.1	23.4	24.2	24.9

 $d=15$ ft. $h=80$ inches

Q	V in feet per second						
	1.5	2.0	2.5	3.0	3.5	4.0	5.0
0.00	0.0	0.0	0.0				
0.04	0.3	0.1	0.0				
0.08	0.7	0.3	0.1	0.0			
0.12	1.3	0.5	0.2	0.1	0.0		
0.16	2.0	1.0	0.4	0.2	0.0		
0.20	3.0	1.5	0.6	0.3	0.1	0.0	
0.24		2.0	0.8	0.4	0.2	0.0	
0.28		2.5	1.1	0.5	0.2	0.0	
0.32		2.9	1.5	0.6	0.3	0.1	0.0
0.36		3.4	1.9	0.7	0.3	0.1	0.0
0.40		3.9	2.3	0.8	0.4	0.1	0.0
0.44		4.4	2.7	1.0	0.4	0.2	0.0
0.48		5.0	3.1	1.1	0.5	0.2	0.0

 $h=60$ inches

Q	V in feet per second						
	1.5	2.0	2.5	3.0	3.5	4.0	5.0
0.00	0.0	0.0	0.0				
0.04	0.7	0.3	0.1	0.0			
0.08	1.7	0.7	0.3	0.1	0.0		
0.12	2.7	1.5	0.7	0.3	0.1	0.0	
0.16	3.8	2.5	1.2	0.5	0.3	0.1	
0.20	4.8	3.5	1.9	0.9	0.5	0.2	0.0
0.24		4.5	2.5	1.4	0.7	0.4	0.1
0.28		5.5	3.4	1.9	1.0	0.6	0.2
0.32		6.2	4.1	2.4	1.4	0.9	0.4
0.36		6.8	4.9	3.1	1.9	1.2	0.6
0.40		7.5	5.8	3.9	2.4	1.6	0.9
0.44		8.2	6.7	4.6	3.0	2.0	1.2
0.48		8.9	7.7	5.6	3.8	2.6	1.6

Tables of θ (in $^{\circ}\text{F.}$) for various values of d (cont.)

(Q in therms/yard.hour)

 $d=15\text{ ft.}$ $h=40\text{ inches}$

Q	V in feet per second							
	1.5	2.0	2.5	3.0	3.5	4.0	4.5	5.0
0.00	0.0	0.0	0.0	0.0				
0.04	1.6	1.2	0.8	0.4				
0.08	2.9	2.3	1.7	1.1				
0.12	4.0	3.1	2.4	1.7	0.9			
0.16	5.0	4.2	3.3	2.6	1.6			
0.20	5.9	5.1	4.2	3.5	2.6			
0.24		6.1	5.1	4.5	3.7			
0.28		6.9	6.1	5.5	4.6	3.9	3.1	2.3
0.32		7.5	7.1	6.5	5.6	4.8	4.0	3.1
0.36		8.2	8.1	7.6	6.5	5.6	4.9	4.1
0.40		8.9	9.1	8.7	7.5	6.5	5.7	4.9
0.44		9.6	10.0	9.9	8.4	7.2	6.5	5.6
0.48		10.2	11.0	11.2	9.3	8.0	7.1	6.4

 $h=20\text{ inches}$

Q	V in feet per second							
	1.5	2.0	2.5	3.0	3.5	4.0	4.5	5.0
0.00	0.0	0.0	0.0	0.0				
0.04	2.5	2.5	2.5	2.4				
0.08	4.0	3.8	3.9	3.9				
0.12	4.9	4.4	4.8	4.9	5.0			
0.16	5.6	5.1	5.5	5.7	5.8			
0.20	6.3	5.9	6.4	6.6	6.6			
0.24		6.7	7.1	7.5	7.5			
0.28		7.5	8.0	8.4	8.5	8.5	8.5	8.5
0.32		8.3	8.9	9.3	9.5	9.5	9.5	9.5
0.36		8.9	9.6	10.1	10.4	10.4	10.4	10.4
0.40		9.5	10.4	11.0	11.2	11.2	11.2	11.2
0.44		10.2	11.1	11.9	12.0	12.0	12.0	12.0
0.48		10.9	12.0	12.7	12.7	12.7	12.7	12.7

 $h=8\text{ inches}$

Q	V in feet per second							
	1.5	2.0	2.5	3.0	3.5	4.0	4.5	5.0
0.00	0.0	0.0	0.0	0.0				
0.04	2.6	2.3	2.4	2.6				
0.08	4.0	3.5	3.7	4.0				
0.12	4.9	4.4	4.6	5.0	5.4			
0.16	5.4	5.1	5.5	5.8	6.1			
0.20	5.6	5.9	6.2	6.5	7.0			
0.24		6.5	6.9	7.2	7.8			
0.28		7.0	7.5	7.9	8.5	9.1	9.7	10.3
0.32		7.5	8.0	8.6	9.3	9.9	10.5	11.1
0.36		8.1	8.6	9.4	10.0	10.5	11.2	11.9
0.40		8.6	9.2	10.0	10.6	11.2	12.0	12.5
0.44		9.1	9.9	10.7	11.4	12.0	12.7	13.4
0.48		9.9	10.9	11.7	12.4	13.2	13.9	14.5

Tables of θ (in $^{\circ}\text{F.}$) for various values of d (cont.)

(Q in therms/yard.hour)

 $d=15$ ft. $h=1.6$ inches

Q	V in feet per second							
	1.5	2.0	2.5	3.0	3.5	4.0	4.5	5.0
0.00	0.0	0.0	0.0	0.0				
0.04	2.0	1.7	1.8	2.0				
0.08	3.0	2.6	2.8	3.1				
0.12	3.4	3.3	3.5	4.0	4.3			
0.16	3.7	4.0	4.1	4.5	5.0			
0.20	4.0	4.5	4.6	5.0	5.5			
0.24		4.9	5.1	5.6	6.1			
0.28		5.3	5.5	6.1	6.7	7.2	7.6	7.9
0.32		5.6	6.0	6.6	7.2	7.8	8.2	8.6
0.36		6.0	6.5	7.1	7.8	8.3	8.7	9.1
0.40		6.5	7.0	7.6	8.4	8.9	9.4	9.9
0.44		6.9	7.5	8.2	8.9	9.5	10.0	10.5
0.48		7.3	8.0	8.8	9.5	10.1	10.7	11.2

Tables of v (in ft/sec.) for various values of d

(Q in therms/yard.hour)

 $d=3$ ft. 9 in. $h=24$ inches

Q	V in feet per second							
	1.5	2.0	2.5	3.0	3.5	4.0	4.5	5.0
0.00	1.40	1.87	2.35	2.80	3.27	3.74	4.20	4.66
0.04	1.45	1.90	2.38	2.84	3.31	3.80	4.26	4.73
0.08	1.49	1.94	2.42	2.87	3.35	3.85	4.32	4.79
0.12	1.54	1.98	2.45	2.90	3.39	3.89	4.36	4.83
0.16	1.59	2.02	2.49	2.93	3.43	3.91	4.39	4.86
0.20	1.63	2.07	2.52	2.96	3.45	3.90	4.37	4.83
0.24		2.12	2.56	2.99	3.45	3.89	4.33	4.79
0.28		2.19	2.59	3.02	3.45	3.88	4.30	4.73
0.32		2.26	2.63	3.05	3.45	3.87	4.27	4.67
0.36		2.34	2.66	3.08	3.45	3.89	4.24	4.60

 $h=10$ inches

Q	V in feet per second							
	1.5	2.0	2.5	3.0	3.5	4.0	4.5	5.0
0.00	1.25	1.68	2.08	2.50	2.89	3.28	3.68	4.08
0.04	1.49	1.83	2.22	2.66	3.06	3.45	3.85	4.23
0.08	1.72	1.98	2.36	2.81	3.22	3.59	3.98	4.36
0.12	1.90	2.14	2.50	2.95	3.37	3.72	4.10	4.48
0.16	2.06	2.29	2.64	3.08	3.50	3.84	4.21	4.57
0.20	2.17	2.44	2.78	3.20	3.61	3.94	4.30	4.63
0.24		2.59	2.92	3.31	3.70	4.02	4.36	4.68
0.28		2.74	3.06	3.41	3.77	4.08	4.40	4.71
0.32		2.89	3.20	3.50	3.82	4.12	4.43	4.73
0.36		3.04	3.34	3.58	3.85	4.14	4.45	4.74

Tables of v (in ft/sec.) for various values of d (cont.) $(Q$ in therms/yard.hour)

$d=3$ ft. 9 in.		$h=4$ inches							
Q	V in feet per second								
	1.5	2.0	2.5	3.0	3.5	4.0	4.5	5.0	
0.00	1.07	1.43	1.78	2.14	2.49	2.86	3.20	3.55	
0.04	1.41	1.73	2.05	2.37	2.71	3.05	3.39	3.71	
0.08	1.64	1.93	2.24	2.54	2.87	3.18	3.51	3.80	
0.12	1.84	2.04	2.36	2.65	2.97	3.26	3.57	3.86	
0.16	2.01	2.12	2.42	2.71	3.02	3.30	3.59	3.88	
0.20	2.12	2.19	2.47	2.75	3.03	3.31	3.60	3.87	
0.24		2.26	2.51	2.77	3.03	3.30	3.56	3.81	
0.28		2.33	2.54	2.79	3.02	3.27	3.51	3.76	
0.32		2.40	2.57	2.80	3.00	3.22	3.46	3.70	
0.36		2.46	2.59	2.80	2.98	3.17	3.39	3.63	

$d=7$ ft. 6 in.		$h=48$ inches							
Q	V in feet per second								
	1.5	2.0	2.5	3.0	3.5	4.0	4.5	5.0	
0.00	1.46	1.96	2.46	2.96	3.44	3.93	4.42	4.90	
0.04	1.48	1.96	2.47	2.97	3.46	3.95	4.45	4.95	
0.08	1.49	1.97	2.48	2.98	3.48	3.97	4.47	4.98	
0.12	1.50	1.98	2.49	2.99	3.49	3.99	4.49	4.99	
0.16	1.51	1.99	2.50	3.00	3.50	4.00	4.50	5.00	
0.20	1.52	2.00	2.49	2.99	3.49	3.99	4.49	4.99	
0.24		2.01	2.48	2.98	3.48	3.97	4.48	4.98	
0.28		2.02	2.47	2.97	3.47	3.95	4.46	4.96	
0.32		2.03	2.46	2.96	3.45	3.93	4.44	4.94	
0.36		2.04	2.45	2.95	3.43	3.91	4.42	4.92	

		$h=24$ inches							
Q	V in feet per second								
	1.5	2.0	2.5	3.0	3.5	4.0	4.5	5.0	
0.00	1.40	1.87	2.35	2.80	3.27	3.74	4.20	4.66	
0.04	1.58	1.97	2.43	2.86	3.32	3.77	4.21	4.66	
0.08	1.75	2.07	2.51	2.92	3.36	3.80	4.22	4.66	
0.12	1.90	2.17	2.59	2.98	3.40	3.83	4.24	4.66	
0.16	2.03	2.27	2.67	3.04	3.44	3.86	4.26	4.65	
0.20	2.15	2.37	2.74	3.10	3.48	3.88	4.27	4.64	
0.24		2.46	2.81	3.16	3.52	3.90	4.27	4.63	
0.28		2.55	2.88	3.22	3.55	3.92	4.27	4.61	
0.32		2.63	2.95	3.27	3.58	3.93	4.27	4.59	
0.36		2.71	3.02	3.33	3.61	3.94	4.26	4.57	

Tables of v (in ft/sec.) for various values of d (cont.) $(Q$ in therms./yard.hour) $d=7$ ft. 6 in. $h=10$ inches

Q	V in feet per second						
	1.5	2.0	2.5	3.0	3.5	4.0	5.0
0.00	1.25	1.68	2.08	2.50	2.89	3.28	4.08
0.04	1.52	1.91	2.24	2.63	3.00	3.40	4.15
0.08	1.73	2.09	2.38	2.74	3.09	3.48	4.19
0.12	1.92	2.23	2.51	2.84	3.17	3.54	4.23
0.16	2.07	2.35	2.62	2.93	3.25	3.60	4.27
0.20	2.20	2.46	2.72	3.01	3.32	3.65	4.30
0.24		2.57	2.81	3.08	3.38	3.70	4.32
0.28		2.67	2.89	3.14	3.43	3.75	4.34
0.32		2.77	2.96	3.20	3.48	3.79	4.36
0.36		2.87	3.02	3.25	3.52	3.83	4.38

 $h=4$ inches

Q	V in feet per second						
	1.5	2.0	2.5	3.0	3.5	4.0	5.0
0.00	1.07	1.43	1.78	2.14	2.49	2.86	3.55
0.04	1.39	1.76	2.10	2.45	2.79	3.13	3.82
0.08	1.63	1.97	2.30	2.62	2.96	3.29	3.95
0.12	1.80	2.10	2.43	2.74	3.06	3.39	4.02
0.16	1.94	2.19	2.52	2.81	3.12	3.43	4.06
0.20	2.04	2.26	2.59	2.86	3.16	3.45	4.06
0.24		2.32	2.64	2.90	3.19	3.47	4.04
0.28		2.38	2.67	2.93	3.21	3.48	4.02
0.32		2.43	2.69	2.95	3.23	3.48	4.01
0.36		2.48	2.71	2.96	3.24	3.48	4.00

 $d=15$ ft. $h=72$ inches

Q	V in feet per second			
	2.0	2.5	3.0	3.5
0.00	2.00	2.50	3.00	3.50
0.12	2.00	2.49	2.99	3.50

 $h=60$ inches

Q	V in feet per second			
	2.0	2.5	3.0	3.5
0.00	1.98	2.48	2.98	3.47
0.12	2.08	2.51	2.99	3.47

 $h=48$ inches

Q	V in feet per second			
	2.0	2.5	3.0	3.5
0.00	1.96	2.46	2.96	3.44
0.12	2.16	2.56	3.01	3.49

 $h=36$ inches

Q	V in feet per second			
	2.0	2.5	3.0	3.5
0.00	1.92	2.42	2.90	3.39
0.12	2.23	2.60	3.00	3.46

 $h=24$ inches

Q	V in feet per second			
	2.0	2.5	3.0	3.5
0.00	1.87	2.35	2.80	3.27
0.12	2.27	2.60	2.97	3.42

 $h=10$ inches

Q	V in feet per second			
	2.0	2.5	3.0	3.5
0.00	1.68	2.08	2.50	2.89
0.12	2.14	2.49	2.91	3.35

 $h=4$ inches

Q	V in feet per second			
	2.0	2.5	3.0	3.5
0.00	1.43	1.78	2.14	2.49
0.12	1.99	2.39	2.84	3.31

Physics and the Atmosphere

By G. M. B. DOBSON

Oxford

*5th Charles Chree Address, delivered 4th November 1949; MS. received
28th November 1949*

I FEEL it a very great honour that you should have asked me to give this year's Charles Chree address. I remember, when he was President of the Royal Meteorological Society, Dr. Chree introducing a speaker with the remark that, for him, all meteorologists were divided into two groups, viz. those who had worked at Kew Observatory and those who had not, and that the speaker, fortunately, belonged to the first group. I am personally doubly fortunate, in that, not only did I work at Kew Observatory for some 18 months immediately after leaving Cambridge, but I worked there while Chree was the Superintendent.

A mathematician of the first rank, Chree also did much for the study of the Earth's magnetic field in organizing accurate measurements and in discussing the results. When the magnetic observatory at Eskdalemuir was started he had much to do with the selection of the site and the organization of the observations. I remember him telling me how, after he had carefully tested the rock around the proposed site for magnetic effect, the Office of Works managed to find just the small amount of magnetic rock which had escaped him with which to build the walls, and they had all to be pulled down again when a few feet high and rebuilt with non-magnetic stone!

He was much interested in the various long period and cyclic changes in the Earth's magnetic or electric field, and when a series of measurements had been running for a long time he was most insistent that nothing should be changed in the method of observing lest it might introduce some discontinuity into the series. To some of his juniors this was sometimes rather irksome when they thought they had detected errors in the method of observation which they could easily put right. Perhaps the right thing to have done would have been to have run the old and improved measurements together for some time for inter-comparison. A remarkable characteristic of Chree was his great preference for results to be expressed by tables or figures rather than by a series of curves, and I remember how, when an assistant brought him a foolscap sheet covered with figures—say the hourly values of the Earth's horizontal magnetic force for a month—he would glance through it and almost immediately pick out any irregularity or some interesting point.

He avoided speculation about causes until he thought that sufficient facts were known by which to test such speculation, and if he thought that the observations indicated a definite connection between two phenomena he would test the reality of such a connection in every way possible before he would be satisfied that the connection was real. One of the last papers which Chree wrote (in 1925) was on the results of his investigation of a possible connection between the amount of ozone in the atmosphere and the disturbance of the Earth's magnetic field. The ozone measurements had only been running for about a

year, but after analysing them in many different ways, Chree concluded that there was an apparent connection between the two phenomena. He ended his paper with the words :

"It is never prudent to accept the results of a single year as final. Events analogous to the breaking of the bank of Monte Carlo do occur occasionally even in Nature. But it seems abundantly clear that the systematic continuation of ozone observations for some time to come is desirable in the interests of terrestrial magnetism" (Chree 1926).

On the more personal side I would stress his kindness to all around, and I well remember how he and his sister used to invite me, though much their junior, to their house. Chree was, of course, an ardent Scotsman, and it was a familiar picture to see him every day, after a frugal meal in his room at the Observatory, walking round the Observatory enclosure with a golf-club and ball until two o'clock. I have heard it said that there were few beggars who knocked at his house and had the astuteness to say that they came from north of the Border, and went away empty!

§ 1

The Charles Chree prize is associated with Geophysics, and it has been suggested to me by the Officers that I should give some account of geophysical work with which I have been associated. This I gladly do, for I feel that it is an opportunity to bring before physicists something of the fascinating work that still remains to be done in studying the Physics of the Atmosphere and to suggest that more of them might think of turning their attention to this subject. Experimental geophysical research differs somewhat from the more usual experimental physics in that, while there is much work to be done in the laboratory, a great deal of the work is observing the actual conditions in the Earth or Atmosphere. An American publication recently suggested that one of the great attractions of research on Cosmic Radiation was that it could be made to take you to all sorts of interesting parts of the Earth. I may say that Geophysics is just as helpful in this matter! (While mentioning Cosmic Radiation, let me remind the atomic physicists that it was workers in Geophysics who presented them with this powerful tool, as evidenced by the older German name of 'Hesssche Strahlung,' which might perhaps with advantage be retained today.) But to return to experimental geophysics: this has both its peculiar difficulties and peculiar attractions—difficulties due to the fact that, as in the case of Astrophysics, one must observe Nature as it presents itself, and one cannot, in general, control the conditions of an experiment as one can in laboratory physics, while one of its great attractions is that much of the work is done out of doors under the open sky. It is recorded of Rutherford that, at a time when experiments were going particularly well, he remarked to a colleague: "I do pity those people who haven't got laboratories to work in". The geophysicist, after a good day's observations in the upper Alps, or even on our little hill above Oxford, feels like remarking: "I do pity those poor physicists who have to work cooped up in their stuffy laboratories".

As I see it today, the great need in Geophysics is for much more experimental work. Theorists we want and, indeed, we already have many of high standing, but they seem to me to be seriously hampered by the absence of facts based on experimental work on which to found, and by which to test, their work. May I give two instances of what I mean and to which I shall refer again. In the



theoretical study of the temperature of the stratosphere, the amount of water vapour was, until a few years ago, entirely unknown, and it was generally assumed that the air would be saturated. On this assumption the absorption and emission of long-wave radiation by water vapour would outweigh that of any other gases which might take part in the radiative equilibrium. When, during the war, it became necessary to measure the humidity of the stratosphere on account of the operational importance of aircraft condensation trails, it was at once found that, far from being saturated with water vapour, the air in the stratosphere was, in fact, very dry. Again, at the present time most theoretical work on the stratosphere assumes that the air there is in radiative equilibrium, but no measurements to find out whether this is so have yet been made, though this should be quite possible and some work on the subject is now in hand.

I believe that there is a growing tendency for some physicists to dislike what I may call 'factory methods' in research, where the work can only be carried out by a large team of workers with elaborate and expensive apparatus and where the individual worker—particularly if he be a junior—may feel that he is merely a cog in the machine and his individuality is to some extent submerged, with consequent loss of interest. While this applies to one important branch of Geophysics just opening up, viz. the exploration of the highest atmosphere by rockets, much work of value can be carried out by individuals without elaborate apparatus. The first pioneer work of Aitken and, later, of C. T. R. Wilson on the condensation of water vapour and the importance of condensation nuclei was carried out with the simplest of apparatus (Aitken 1880, 1887, Wilson 1895, 1900). When W. H. Dines started to measure the temperature of the upper atmosphere by instruments carried on free balloons, the available funds were very small, but by very skilful design he was able to make instruments which were so simple that they could easily be made in quantity in his workshop, and so light that they needed only small, and therefore cheap, balloons to lift them. With one instrument maker and a very small grant from the Meteorological Office he was largely responsible for first obtaining most of our knowledge of the temperature and pressure of the upper atmosphere and the correlations between them (Dines 1909 to 1919). When in the early years of this century Mr. C. J. P. Cave began to study the wind at various heights above the Earth's surface, much of the work was carried out with the sole help of his gardener for launching the balloons (Cave 1912).

There are still many problems in Geophysics waiting to be solved which can well be tackled by individual workers with relatively simple apparatus. Any physicist who is prepared to be his own instrument maker will find a wide field open to him.

§ 2. ATMOSPHERIC OZONE

Turning now to describe some of the geophysical work that has been carried out at Oxford, I will first take the measurements of the amount of atmospheric ozone. The total amount of ozone in the air is less than a millionth of the oxygen and nitrogen, and this small amount is very largely situated high up in the atmosphere, chiefly between 10 and 30 km. Anyone might therefore well ask why it should be of much interest to study the variations of such a rare constituent of the atmosphere. There are several things which make this small

amount of ozone very interesting. As you will know, ozone very strongly absorbs ultra-violet radiation of wavelength shorter than 3200 Å. and cuts off all the solar radiation in shorter wavelengths so that, until the advent of high flying rockets, the solar spectrum was entirely unknown for wavelengths shorter than 2900 Å. Secondly, as Dr. Gowan has shown (Gowan 1936), the upper part of the ozone region absorbs enough solar radiation to cause the temperature at heights of 40 to 60 km. to rise to values above that at the surface of the Earth at the Equator. (Incidentally, the presence of this warm region is responsible for a number of other interesting effects which I have not time to discuss here.) Thirdly, there is some evidence that the energy of the outgoing terrestrial radiation of wavelengths around 9μ , which is absorbed by ozone, is sufficient to have a marked effect on the temperature of the stratosphere. Finally, it is an observed fact that the amount of ozone in the upper atmosphere is closely correlated with the general meteorological conditions and particularly with the conditions at great heights. It may well be that when our knowledge of the subject is more complete we shall find such measurements of practical importance.

The method used to measure the ozone content of the atmosphere follows closely that employed by Fabry and Buisson (1913, 1921), who were the first to make any accurate measurements of atmospheric ozone. Utilizing the strong ultra-violet absorption band of ozone, which increases in intensity from about 3300 Å. to a maximum at about 2500 Å., measurements of the relative intensity of the solar spectrum at two wavelengths on the long-wave edge of this absorption band are used to deduce the amount of ozone through which the light has passed and hence the ozone content of the atmosphere above the place of observation. In designing the apparatus for such measurements great care must be taken to avoid errors due to light scattered within the spectrograph since the intensity of the Sun's radiation at these short wavelengths after absorption by ozone is extremely small compared with that in the visible region.

Simple Feré spectrographs built in my laboratory were used up to about 1930 and gave excellent results when direct sunlight was available, but they suffered from the great disadvantage that the photographic plates had to be sent back for development and measurement, so that the ozone values were not known until several days after the spectra were taken, while no measurements were possible on completely cloudy days. Largely for this reason, a much more elaborate photoelectric spectrophotometer was built which allowed the ozone content to be measured by the absorption of ultra-violet light in the same way as with the photographic instruments. The latter instrument is both more sensitive and also allows the ozone value to be calculated immediately after the observation, observation and calculation taking about ten minutes. The first instrument was built in my laboratory, but when a large number were required they had to be manufactured by an optical firm. Some nineteen of those photoelectric instruments had been made at the beginning of the war and some fourteen more instruments are now under construction.

While the pre-war instruments used sodium photoelectric cells, the present instruments use photo-multipliers. Photo-multipliers have the great advantage of low 'signal/noise ratio' and hence allow measurements to be made with much weaker light than was possible before. This is particularly important where the vertical distribution of the ozone in the atmosphere is to be measured by

observations using the light from the zenith sky when the Sun is very low. We can also now make measurements at night using the light of the full moon with fairly good accuracy, and such measurements should be valuable in very high latitudes in winter.

It is possible to estimate the vertical distribution of the atmospheric ozone by measurements of the zenith sky-light and also, more directly, by sending up spectrographs on free balloons or rockets to photograph the solar ultra-violet spectrum at great heights. Both these methods agree in showing that most of the ozone is situated between 10 km. and 30 km. Figure 1 shows the vertical distribution of ozone deduced from observations of the zenith sky-light, while Figure 2 shows spectra obtained at various heights from a rocket in America.

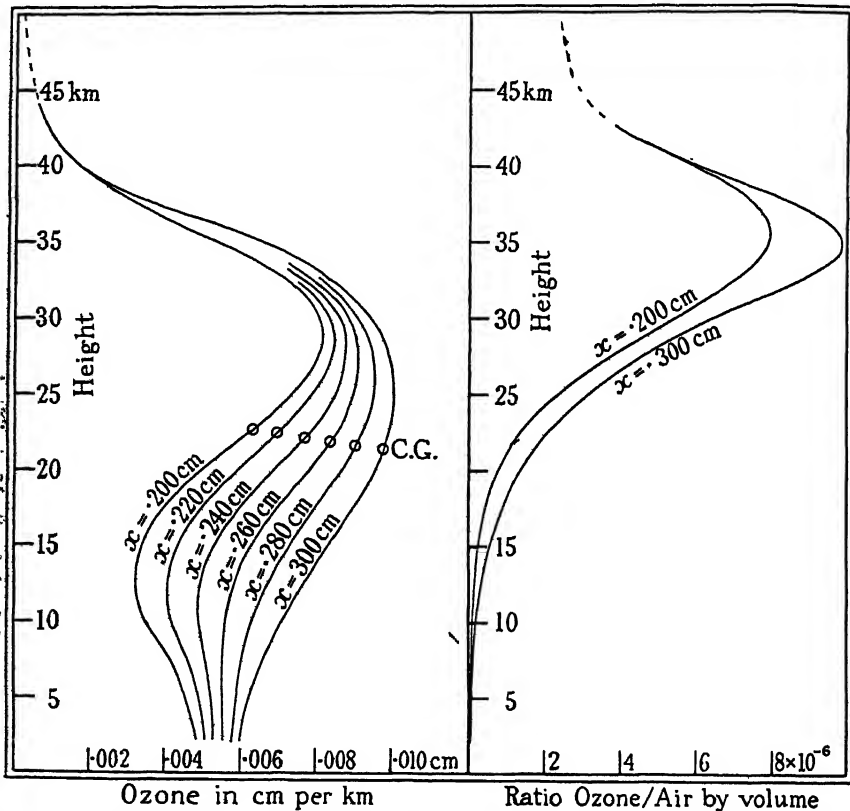


Figure 1. Vertical distribution of ozone as measured at Arosa by observations of the zenith skylight. (*Proc. Roy. Soc. A*, 1943, 145, 440, Figure 8.)

These spectra show, for the first time, the spectrum of the Sun in the region beyond 2900 Å. Spectra obtained from rockets or even high altitude balloons can only be obtained occasionally but, given good weather, the vertical distribution can be deduced from zenith sky-light observations every day and can be used for comparisons with the weather conditions. Contrary to what I said in the early part of this address, in this particular case, while the experimental geophysicist can make the necessary observations with good accuracy, he is at present somewhat held up by lack of theoretical work which would help him to deduce the vertical distribution from those observations.

Figure 3 shows the annual variation of ozone content at eleven places widely distributed over the globe and refers to the years 1926 to 1929. The chief points of interest in this figure are firstly the peculiar annual variation in the ozone content with a maximum in the local spring and minimum in the local autumn for both hemispheres, with little annual variation in low latitudes. Secondly, it is of interest to note the general increase in ozone content from low to high latitudes, this increase being much more marked in the spring than in the autumn.

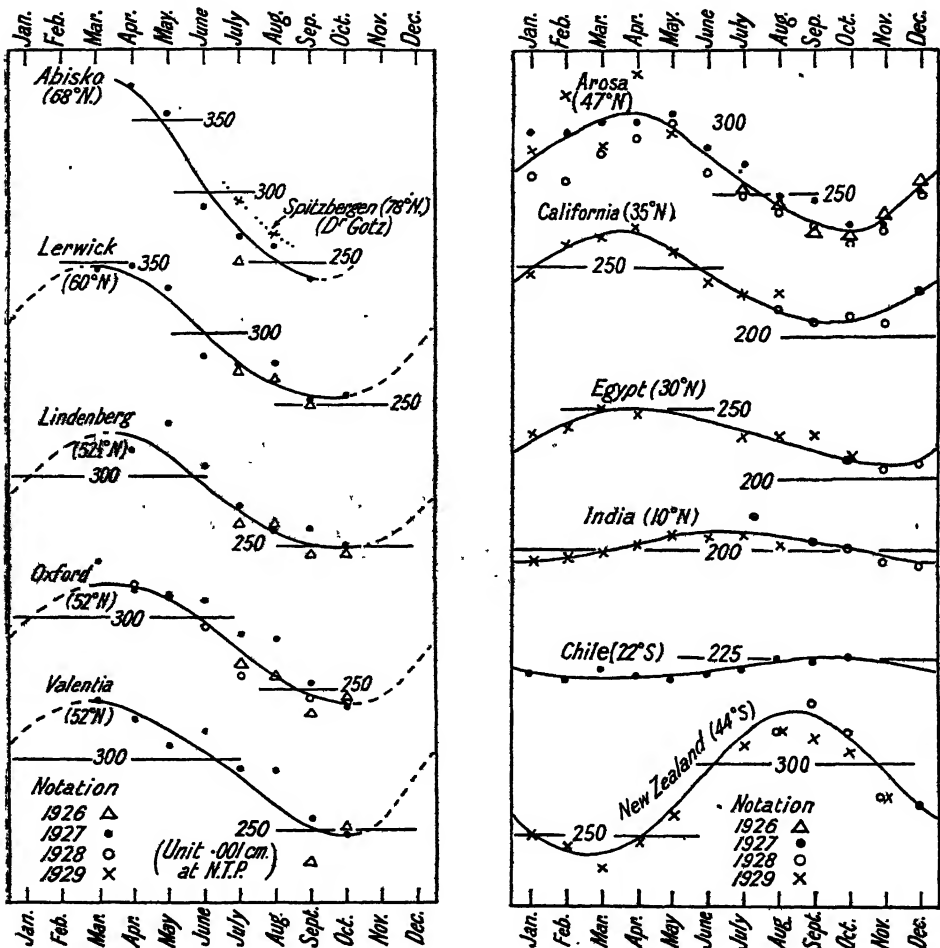
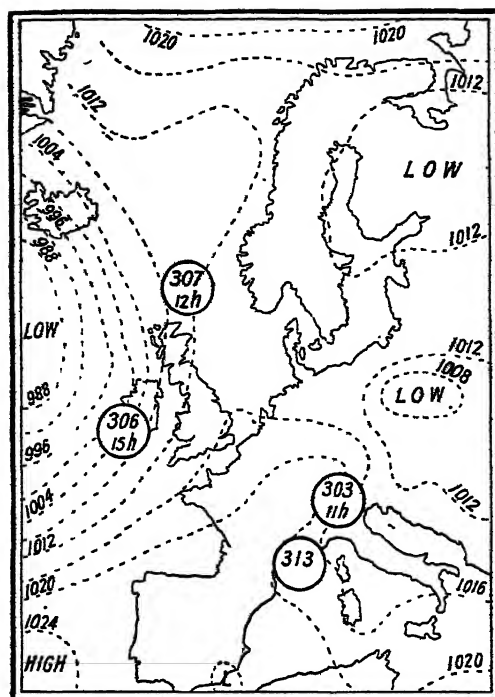


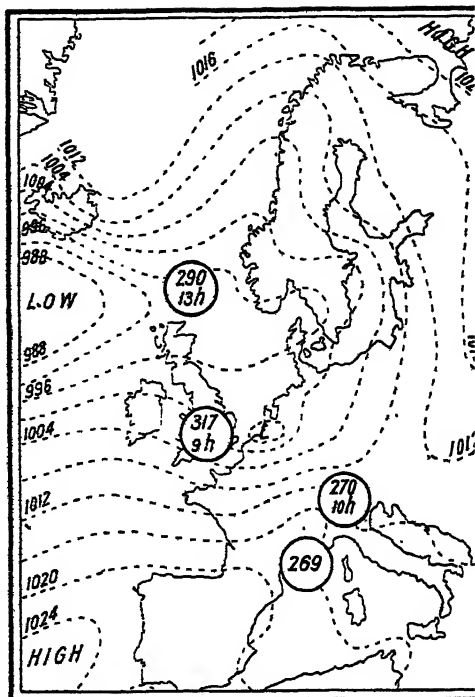
Figure 3. Annual variation of the amount of ozone in the atmosphere in different parts of the world. (*Proc. Roy. Soc. A*, 1930, 129, 417, Figures 1 (a) and 1 (b).)

As an illustration of the way in which the amount of ozone varies from day to day with weather conditions, Figure 4 shows the distribution of ozone over N.W. Europe on 4th, 6th, 7th and 8th April 1927, when a depression, with an associated secondary, passed across the British Isles. I would particularly draw your attention to the large increase in atmospheric ozone in S.W. Ireland (0.306 cm. to 0.420 cm.) as the secondary passed over. On the other hand, Figure 5 shows the large fall in ozone content, first at S.W. Ireland and later in Great Britain, as a warm front approached from the west.



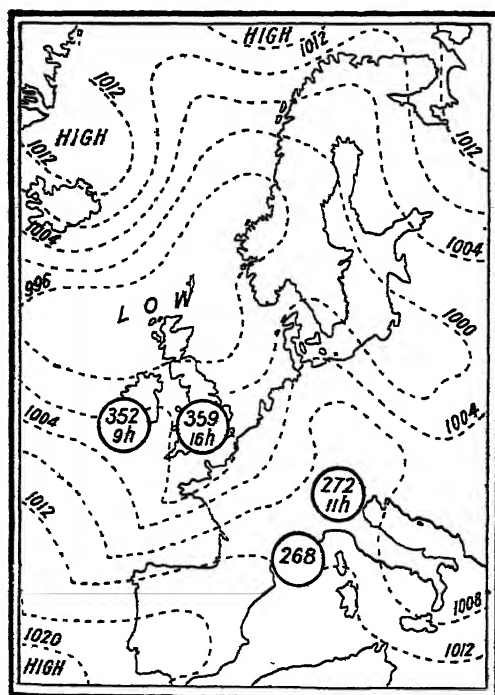
April 4, 1927.

7 A.M.



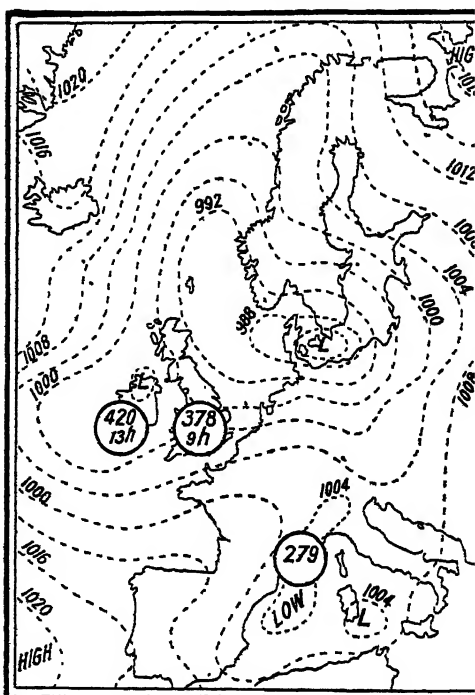
April 6, 1927.

7 A.M.



April 7, 1927.

7 A.M.



April 8, 1927.

7 A.M.

Figure 4. Figures within the circles show the amount of ozone (unit 0.001 cm.) at various places N.W. Europe. Note the large increase in the amount of ozone in the secondary depression on 8th Apr (*Proc. Roy. Soc. A*, 1929, 122, 473, Figure 5.)

As a further example of my previous point, that much work in Geophysics can be done with very simple apparatus, I would say that the marked connection between the ozone content and the weather, shown more fully in Figures 4 and 5, was first indicated by the results given by a simple Feré spectrograph built in my small laboratory for a total cost of under £25, while the observations at the various stations shown in Figures 3 and 4 were made by similar instruments also built in my small laboratory, though with a grant from the Royal Society for the optical parts.

The changes in atmospheric ozone which occur between one day and the next in temperate latitudes are almost certainly due, in part, to large-scale transport of air from higher or from lower latitudes, but whether this is the whole cause is by no means certain at the present time. A co-operative scheme is now being organized by Sir Charles Normand, under the auspices of the International

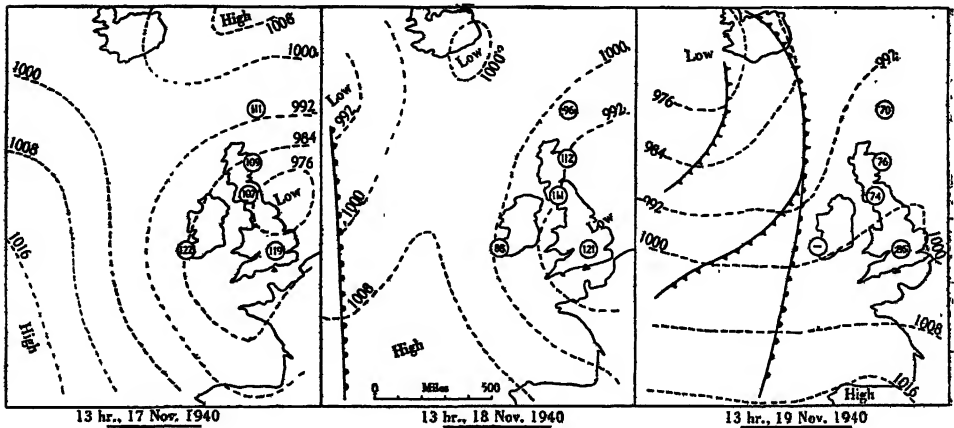


Figure 5. Figures within the circle give the amount of ozone at five places in the British Isles expressed as percentages of the mean value for the season. Note the large fall in the west of Ireland on 18th November and in Great Britain on 19th November, when the warm front, which caused the fall, was still several hundred miles to the west.

(*Proc. Roy. Soc. A*, 1946, 185, 160, Figure 8.)

Meteorological Association and with the generous support of the Royal Society, to study synoptically the changes in the ozone content at a dozen or more stations distributed over N.W. Europe, and to relate these changes to the meteorological conditions. This involves very careful inter-comparison at Oxford of all the instruments before they are distributed, so that the ozone values obtained from them shall be comparable. Several other instruments destined for more distant parts of the world are also being calibrated at Oxford.

The earlier synoptic studies from which Figures 4 and 5 are taken suffered from a paucity of ozone observing stations, but quite as much from a paucity of observations of the meteorological conditions in the upper atmosphere. At the present time the number of daily meteorological balloon ascents has been greatly increased and should be adequate for such a study, though we may perhaps wish for meteorological observations at even greater heights than those now available since the correlation between the zone changes and other meteorological conditions appears to become closer with increasing height up to the limit reached by meteorological balloons.

§3. THE HUMIDITY OF THE UPPER ATMOSPHERE

For many years measurements of the humidity of the upper air have been made by wet and dry bulb psychrometers on aircraft and by small instruments carried up by balloons. In this country the instruments carried by balloons used the expansion of hairs or gold-beater's skin to measure the relative humidity, while in America the changes of electrical resistance of a thin film containing a hygroscopic salt was employed. Unfortunately all these methods become impracticable at very low temperatures where the amount of water vapour is very small. After a thorough examination of all the possible methods of measuring humidity, we concluded in 1941 that the dew-point (or rather frost-point) method was the most hopeful for use at very low temperatures. The first problem to solve was how to measure the very small amount of deposit of dew or hoar-frost on the thimble of the hygrometer. The driest air which has so far been found in the atmosphere contains only about 300 micrograms of water vapour per cubic metre. Since only a few litres of air can be passed over the hygrometer thimble in a minute, and since only a very small fraction of the water vapour can be removed from the air, it will be seen that we have to measure an amount of water vapour of the order of a few micrograms. If this were spread uniformly over the thimble it would be quite invisible. Fortunately, with the exception to be noted later, dew and hoar-frost tend to be deposited as droplets or particles of a size very suitable for scattering light, and if the clean thimble surface scatters very little light, it is possible to see extremely small amounts of deposit. The amount of deposit may be judged by eye observation or may be measured more accurately by the aid of a photocell. In the first case the air stream is made to impinge on to the thimble as a narrow jet so that the deposit tends to form as a streak across the otherwise clean thimble surface. The eye is thus helped by the contrast between the clean thimble and the deposit. In the second case the thimble is illuminated by a strong beam of light inclined at about 45° to the surface, while the photocell receives only light scattered normally from the surface. If the polishing of the thimble surface is always done in the plane of the beam of light, very little light is scattered by the clean thimble. Using a single valve amplification, it is easy to detect extremely small deposits of hoar-frost on the thimble. When using the eye-observation method it is necessary to find two temperatures of the thimble (as close together as possible) at which the deposit is firstly slowly growing and secondly equally slowly evaporating. The photoelectric method is so much more sensitive that one has only to cool the thimble until a small deposit has formed on it and then raise the temperature until the indicator shows that the amount of deposit is remaining constant.

Using such methods one can measure the frost-point of air down to temperatures of about 190°K. (-83°C.), where the water-vapour content of saturated air is only about 300 micrograms per cubic metre. Unfortunately at still lower temperatures ice tends to be deposited as a glassy layer which does not scatter light. It is possible, in principle, to measure changes in the thickness of the layer of glassy ice by optical interference methods, but as the layer would only change in thickness by one-tenth to one-hundredth of a wavelength per minute, one could not expect to measure air with frost-points much below those which have been measured with existing instruments.

In this work ceaseless watch must be kept for possible errors, one of the chief being contamination of the air on its way to the hygrometer by water vapour

given off from the walls of the piping. To avoid errors of this kind, a clean metal tube, at least $\frac{1}{2}$ inch in diameter, is led from outside the aircraft to the hygrometer and out again to the outside air. As the entrance of the tube faces forward, a strong blast of air passes through the tube continually when the aircraft is in flight. A very small fraction of this air stream is led by a very short branch tube to the thimble of the hygrometer. A cap covers the external entrance to the tube when the aircraft is flying through cloud.

There had been much speculation among meteorologists about the humidity of the air in the stratosphere, and it was therefore with great interest that the first results were awaited. The first accurate measurements were made by Mr. Brewer in a Fortress aircraft. On the 26th August and the 7th September

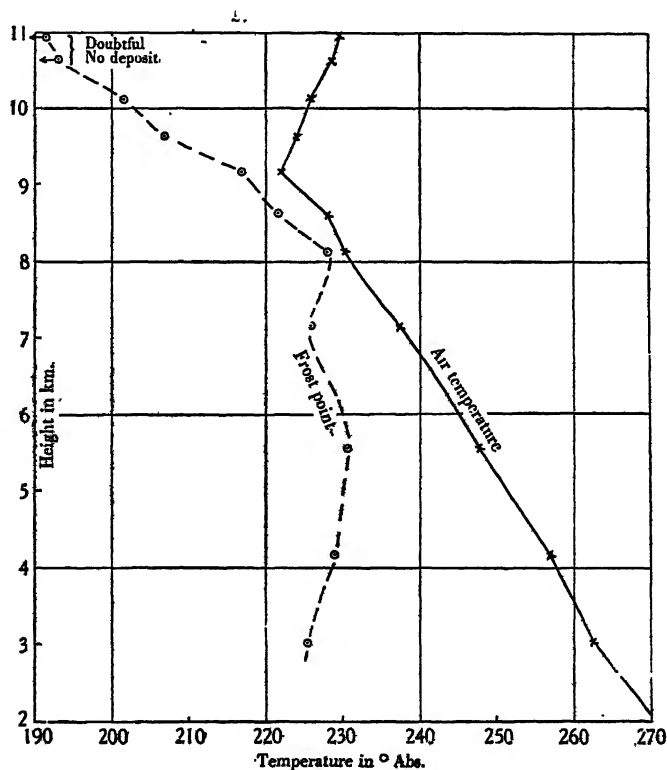


Figure 6. Frost-points and temperatures of the upper air on 22nd December 1943, this being the first case of a successful measurement of the humidity in the stratosphere. Note the very dry air in the stratosphere, indicated by the large difference between frost-point and air temperature.

(*Proc. Roy. Soc. A*, 1946, 185, 154, Figure 4.)

1943 the aircraft just entered the stratosphere and the measurements at the top of the climb indicated a great reduction in humidity. On the 22nd December 1943 the aircraft reached heights well within the stratosphere and the measurements left no doubt that the stratosphere air was very dry (see Figure 6). This conclusion has been confirmed by some hundred or more ascents which have since been made by the Meteorological Research Flight of the Air Ministry. The early measurements were made in aircraft without pressure cabins, and success was as much due to physical endurance as to observational skill. Even now no measurements have been made of the stratosphere when the tropopause is very high.

In Figure 6 it will be noted that the rapid fall in frost-point with height begins about one kilometre below the tropopause and continues up to the maximum height attained. In Figure 7 the troposphere was nearly saturated throughout and there was a marked increase in the rate of fall of the frost-point immediately the stratosphere was entered. In some cases the rapid fall in the frost-point seems to begin a little above the tropopause and in others a little below it. Figure 8 shows results given by Mr. Shellard (Shellard 1949) for the average rate of change of both temperature and frost-point in the neighbourhood of the tropopause, in summer, in winter and for the year. In this diagram all heights are reckoned upwards and downwards from the tropopause. It will be seen that there is an increase in the rate of fall of the frost-point immediately above the tropopause, but that the rate of fall decreases again near the maximum height reached, though there is no evidence of the frost-point becoming constant or increasing again at greater heights. It is a question of much interest to know whether the air becomes still drier above the highest aircraft observations.

In addition to the very dry air found in the stratosphere, the frost-point hygrometer also frequently shows most interesting shallow dry layers in the troposphere associated with temperature inversions. It is remarkable that these shallow dry layers should be able to persist between much damper layers above and below them, and it is by no means clear at present where the dry air came from.

The origin of the very dry air in the stratosphere has also given rise to much speculation. The only obvious way in which the air could have been dried is by having been cooled to a low temperature, when the water vapour would be condensed and fall out of the air. The lowest natural temperature ever measured in the atmosphere is a little below 190°K . (-83°C .) in the air at about 20 km. over the Equator. The lowest measured frost-points are almost down to this figure, and one can therefore just explain them by the hypothesis that the air was dried over the Equator and has not since picked up any appreciable amount of water vapour. It should be noticed, however, that the wind currents in the stratosphere may come from any direction and the air is always very dry. In many cases it must have been many days or weeks since the air was last in the region of the low temperature over the Equator.

Very interesting work is going on at Cambridge with the object of measuring the intensity of sunlight in the region of the $6.3\ \mu$ absorption band of water vapour at great heights from aircraft. Such measurements should indicate the total amount of water vapour above the height of the aircraft, and from them it should be possible to deduce something about the humidity of the air at heights where at present we cannot operate frost-point hygrometers. It is also hoped that it may be possible shortly to develop frost-point hygrometers suitable for use on free balloons. Not only will the balloons reach greater heights than can be attained by aircraft, but it should be possible to send up several instruments daily and so study the distribution of humidity synoptically.

§ 4. ICE-FORMING NUCLEI IN THE ATMOSPHERE

A short time ago I witnessed the following phenomenon: a bank of alto-cumulus cloud, consisting of rather large dense cloudlets, was moving eastwards across the sky and, when each cloudlet reached a given position, snow began to fall from it and the cloudlet vanished in a few minutes, only a faint trail

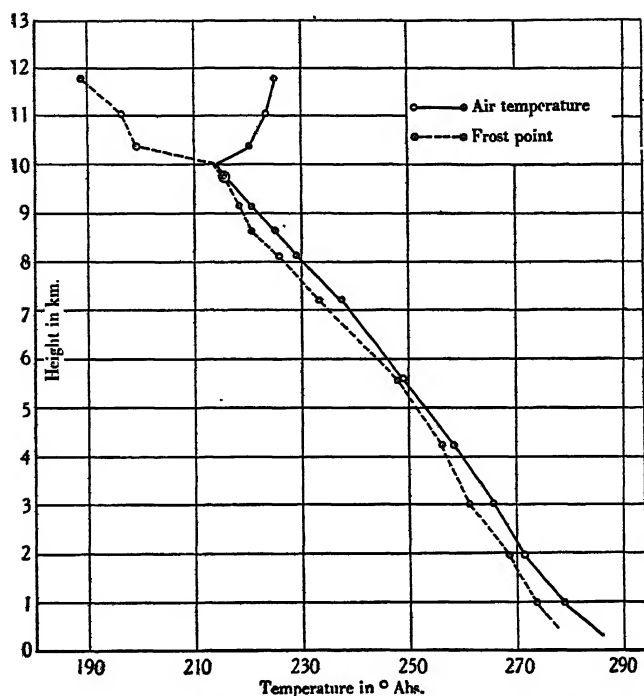


Figure 7. Frost-points and temperature of the upper air on 30th May 1945. Note that in this ascent the air throughout the troposphere is nearly saturated but there is a sharp increase in the rate of fall of the frost-point on entering the stratosphere.

(*Proc. Roy. Soc. A*, 1946, 185, 156, Figure 5 (a).)

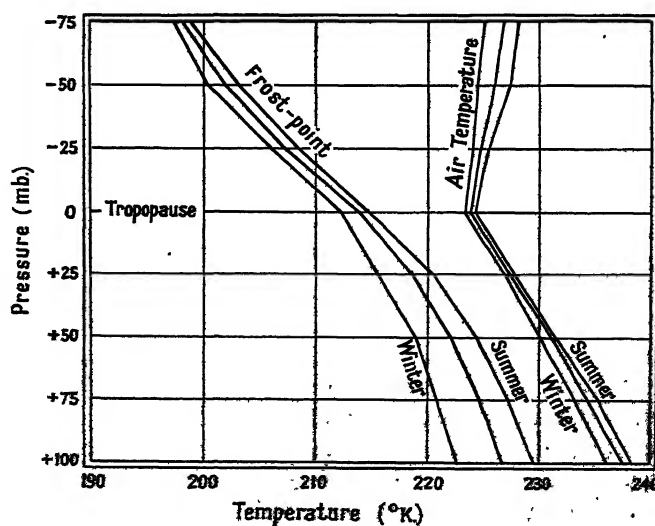


Figure 8. Average change of frost-point and air temperature with height in the neighbourhood of the tropopause.

(Reproduced from *Meteorological Research Paper* 486 by H. C. Shellard, by permission of the Director of the Meteorological Office.)

of falling snow remaining. A striking laboratory demonstration is easily made showing a similar result. A supercooled water cloud is formed by breathing into a refrigerated chamber at, say, 260°K. (-13°C.) and then, if a piece of solid CO_2 be passed through the cloud, it will leave behind many minute crystals of ice. Within a few seconds the water droplets have evaporated and the vapour condensed on the ice crystals, which quickly fall out if they are not too numerous and, therefore, too small.

The history of the alto-cumulus cloud droplets and the water fog formed in the cold vessel illustrate a problem about which I now wish to speak briefly. The formation of ice crystals in clouds is also of much interest since Professor Berg  ron has suggested that all raindrops, except those in a drizzle, first start their existence as ice crystals, which, because of their lower vapour pressure compared with water drops, are able to grow large. Work of great interest has recently been done at Oxford by Cwilog, Fournier d'Albe, Palmer and Brewer on ice-forming nuclei, while similar work has been carried on independently in Germany by Findeisen and by Weickmann and in America by Langmuir and his associates.

Until recently it was thought that ice crystals would be formed on special 'sublimation nuclei' when the air was saturated with respect to ice in the same way as water droplets are formed on 'condensation nuclei' when the air is saturated with respect to water. The first result of the work was to show that if 'sublimation nuclei' existed at all they were exceedingly rare, and we are now able to show that true 'sublimation nuclei' probably do not exist at all in the atmosphere. The work at Oxford has shown that while there are no 'sublimation nuclei' there are many 'ice-forming nuclei,' and that each nucleus has a definite threshold temperature below which it begins to act as an 'ice-forming nucleus'. While the number of nuclei which are active at any particular temperature increases as the temperature is lowered, there are two special threshold temperatures at which large numbers of particles become active as ice-forming nuclei. The nuclei can therefore be divided into three groups. The first group consists of very few nuclei (though they may be very important meteorologically) which are active at temperatures above 241°K. (-32°C.). The number which are active at temperatures as high as 258°K. (-15°C.) is probably negligible, but as the temperature falls to 241°K. (-32°C.) the number of active nuclei gradually increases.

When the temperature reaches 241°K. (-32°C.) a large number of particles are normally found in the lower air over land which become active as 'ice-forming nuclei', and these belong to the second group. When the temperature falls to 232°K. (-41°C.) all particles of whatever kind which may be floating in the air seem to become active ice-forming nuclei. This threshold temperature is very sharp indeed. Particles, as different as sodium chloride or the dust produced by sparks between metal electrodes, all act exactly alike in this way. Moreover they all have ice formed on them only when the humidity of the air reaches (or very closely approaches) saturation over *water*. It now seems (in contradiction to some earlier experiments) that in perfectly clean air no water droplets or ice particles are formed at any temperature unless the air is very highly supersaturated.

It is now easy to explain the formation of the water fog in the cold chamber and its subsequent transformation into an ice fog. There being no 'ice-forming nuclei' which are active at 260°K. (-13°C.), a water fog is formed and, since there are very many condensation nuclei present in surface air, a very large number

of small droplets will be formed which will settle very slowly, and the cloud will be stable. When we draw a piece of solid CO_2 through the water fog the air in immediate contact with the CO_2 will be cooled to a temperature below 232°K. (-41°C.), so that many particles will be active as ice-forming nuclei and a trail of ice crystals will be left in its track. These will be in air supersaturated with respect to ice so that they will grow rapidly and fall out, so that the fog will disappear. (If a very large number of ice particles are formed they may remain small and a stable ice fog will be formed.)

We can also offer a reasonable explanation of why the alto-cumulus clouds persisted for a long time as water clouds and then were quickly transformed into large snowflakes. At the height of the alto-cumulus there would probably be very few, if any, particles which were active ice-forming nuclei at the temperature of the cloud. Hence the cloud formed and continued as a water cloud. I presume that at the later stage the cloudlets drifted into a region where ice crystals were falling from some higher cloud—probably too thin to be seen—there being probably a change of wind with height above the alto-cumulus cloud. The water-drops evaporated in the presence of the ice particles, just as in the experiment in the cold chamber, owing to the higher vapour pressure over water than over ice, and the ice particles grew rapidly for the same reason.

Cwiling made measurements at the Jungfraujoch and over the oceans and showed that in these conditions ice-forming nuclei which were active at temperatures above 232°K. (-41°C.) were generally absent or very rare. Palmer has used the same apparatus in aircraft and has shown that these nuclei are also absent (or very rare) above the level of the first temperature inversion or a low cloud layer, though there are, of course, many condensation nuclei present which will also act as 'ice-forming nuclei' at temperatures below 232°K. (-41°C.). Whether there are *any* of the ice-forming nuclei active at temperatures above 232°K. in the upper atmosphere is not known, and it is extremely difficult to detect the presence of very rare nuclei.

Findeisen observed that the actual freezing of water or sublimation of water vapour seemed to throw off nuclei on which ice can form (possibly they are actual spicules of ice). The production of 'ice-forming nuclei' by these processes has been confirmed at Oxford, and the matter is still under investigation. If the formation of these new nuclei is rapid enough, it is possible that a very few ice crystals, formed by some means in the upper part of a supercooled water cloud, could throw off more nuclei as they grow, and on these nuclei other ice particles would be formed. In this way a very few ice-forming nuclei might lead to the whole cloud of small water droplets evaporating and the vapour condensing on ice particles which would grow very large and fall rapidly. As mentioned before, very little is at present known about this process and it may or may not be important in the formation of rain and snow in the atmosphere.

I have been able to describe only three out of many problems in Geophysics which are awaiting investigation, but I hope that I have shown both that there is much to do which is of great interest and also that elaborate and expensive apparatus is not required. Many of the problems are, as I have said, suitable for individual workers, or a small group, in one of the smaller physical laboratories where such things as cyclotrons and million-volt generators are out of the question. There have never been more than three persons at any one time actively engaged on any one of the questions that I have discussed.

Finally I would thank you again as Members of the Physical Society for doing me the great honour of allowing my name to be associated in this way with that of Charles Chree.

ACKNOWLEDGMENTS

Figures 1, 3, 4, 5, 6 and 7 are reproduced from the *Proceedings of the Royal Society* by permission of the Council. Figure 8 is reproduced by permission of the Director of the Meteorological Office. Figure 2 is reproduced from *Sky and Telescope*.

REFERENCES

- AITKEN, J., 1880, *Trans. Roy. Soc. Edin.*, 30; 1887, *Ibid.*, 33.
 BREWER, A. W., and PALMER, H. P., 1949, *Nature, Lond.*, 164, 312.
 CAVE, C. J. P., 1912, *The Structure of the Atmosphere in Clear Weather* (Cambridge : University Press).
 CHREE, C., 1926, *Proc. Roy. Soc. A*, 110, 693.
 CWILONG, B. M., 1947, *Proc. Roy. Soc. A*, 190, 137.
 DINES, W. H., 1909, *Met. Office Publication* No. 202; 1911, *Phil. Trans. Roy. Soc. A*, 211; 1912, *Met. Office, Geophys. Memoir* No. 2; 1919, *Ibid.*, No. 13.
 FABRY, CH., and BUISSON, H., 1913, *J. Phys. Radium*, ser. 5, 3, 196; 1921, *Ibid.*, ser. 6 2, 197.
 FOURNIER D'ALBE, E. M., 1949, *Quart. J. R. Met. Soc.*, 75, 1.
 GOWAN, E. H., 1936, *Quart. J. R. Met. Soc.*, 62, 34.
 PALMER, H. P., 1949, *Quart. J. R. Met. Soc.*, 75, 15.
 SHELLARD, H. C., 1949, *Meteorological Research Paper* No. 486 (London : Air Ministry).
 WILSON, C. T. R., 1895-1900, *Proc. Camb. Phil. Soc.* and *Proc. Roy. Soc. A*, sundry papers.

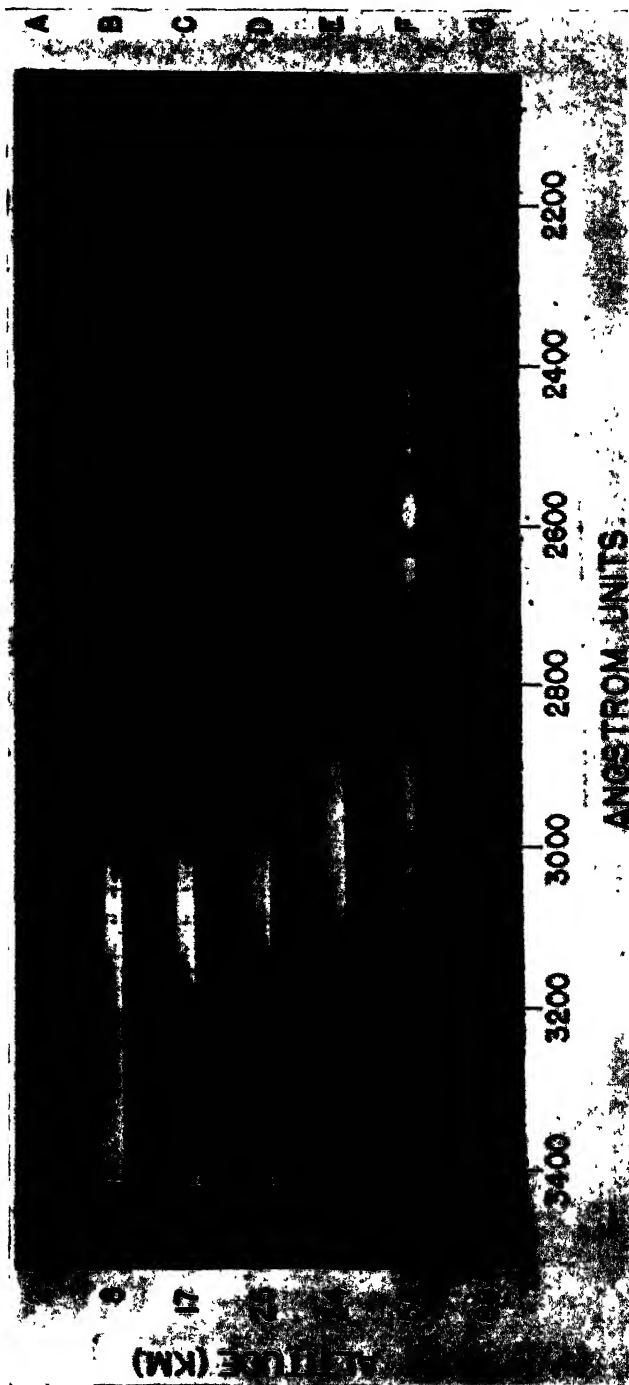


Figure 2. Spectra of solar radiation obtained at different heights by a V.2 rocket sent up in America. Note that in the last two spectra the rocket had rolled over and the spectrograph was not properly exposed.
(From *Sky and Telescope*.)

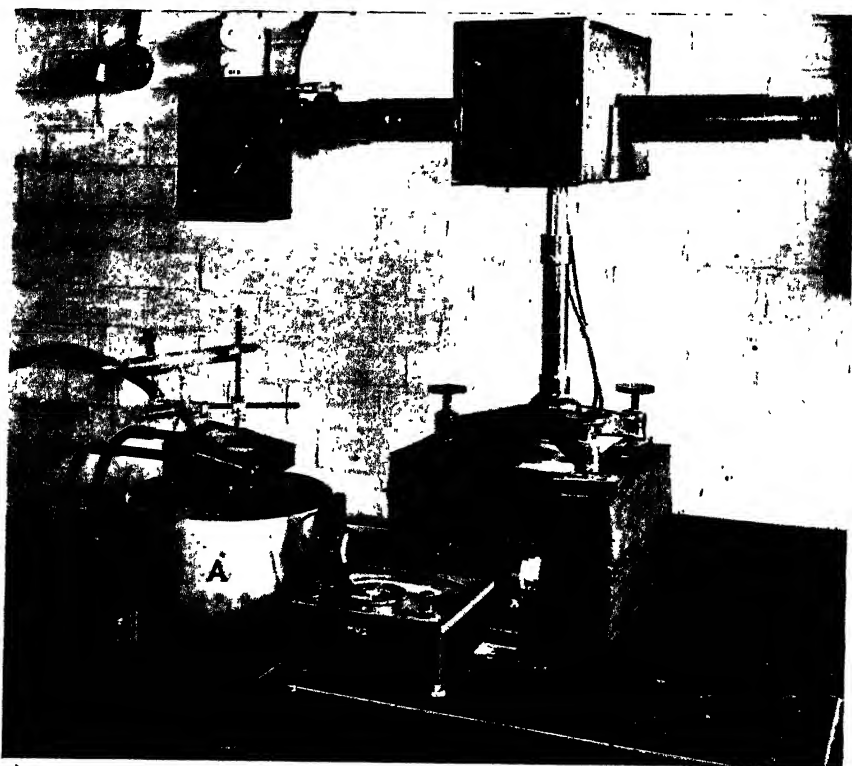


Figure 2. The assembled dilatometer.



Figure 4. Fringes from expansion interferometer (above) and interference thermometer (below).

The Measurement of the Thermal Expansion of Single Crystals of Tin by an Interferometric Method

By B. G. CHILDS AND S. WEINTROUB

University College, Southampton

MS. received 27th September 1949, and in amended form 14th November 1949

ABSTRACT. Single crystals of tin have been grown by the Kapitza method, their orientations measured, and the crystals converted into sets of three spacers for use in an interferometric dilatometer. The dilatometer consisted of a vacuum furnace for heating the spacers and interferometer, and an auto-collimating monochromator. The movement of the fringes with change of temperature was observed visually. The apparatus, its assembly, and method of use are described in detail.

The linear expansion coefficients, α_θ , of six crystals of different orientations, θ , have been measured over the range of temperature 30–220° C., and α_θ for a given temperature was found to vary linearly with $\cos^2 \theta$ in accordance with Voigt's relation. The two principal coefficients, $\alpha_{||}$ and α_{\perp} , corresponding to $\theta=0$ and $\theta=90^\circ$ were found by extrapolation. $\alpha_{||}$ and α_{\perp} increase with temperature from 32.0 and $16.2 \times 10^{-6}/\text{deg. C.}$ at 30° C. to 41.5 and $20.3 \times 10^{-6}/\text{deg. C.}$ at 230° C. respectively. The accuracy of the determination is estimated to be within 1%. The results are exhibited in graphical and tabular form.

§ 1. INTRODUCTION

FOR metals with tetragonal structure, such as tin above its transition temperature, the linear thermal expansion coefficient, α , is not independent of the direction in the lattice as is the case for isotropic metals. The expansion of the lattice in any specified direction can however be calculated from a knowledge of two independent coefficients $\alpha_{||}$ and α_{\perp} which refer respectively to the directions parallel and perpendicular to the principal tetragonal axis. Thus, for a direction making angle θ with the axis, the corresponding expansion coefficient α_θ is, in accordance with the relation stated by Voigt (1910),

$$\alpha_\theta = \alpha_{||} \cos^2 \theta + \alpha_{\perp} \sin^2 \theta.$$

Since the coefficients vary considerably with temperature, a complete description of the thermal expansion of tetragonal tin must consist of values of $\alpha_{||}$ and α_{\perp} for all temperatures between the transition point (13° C.) and the melting point (232° C.).

In the present paper a description is given of measurements, by an interferometric method, of the expansion of tin single crystals, of various orientations, over the range from room temperature up to the melting point. It is believed that this is the first time that such an investigation has been made. The results, which are discussed in detail in § 7, show that Voigt's relation is obeyed, that $\alpha_{||}$ is approximately twice α_{\perp} , and that both $\alpha_{||}$ and α_{\perp} steadily increase with temperature over the whole range of temperature.

§ 2. METHOD OF MEASUREMENT

Of the several methods which have been used for the measurement of the linear expansion of small crystal specimens, that which is capable of the greatest accuracy is the Fizeau (1869) interferometric method. The specimen in this

case is converted into a form suitable for the three point support of an optically flat, glass or quartz, plate vertically above another similar plate, the whole assembly forming an optical interferometer. If the three pieces of the specimen, or 'spacers', are nearly equal in length, so that the wedge angle between the inner plate surfaces is not greater than some four seconds of arc, a fringe system is produced by interference of the reflections, at the interior surfaces of the plates, of a parallel beam of monochromatic light of wavelength λ , incident normally on the plates. The fringe system, which is located in the 'film', the space between the plates, consists of contour lines of equal film thickness at intervals of $\lambda/2$.

When the temperature of the interferometer is raised, thus causing the spacers to expand, the fringes move laterally, and their movement past a fixed point serves to measure the expansion of the film at that point. Under the most favourable conditions using automatic recording it may be possible to observe the fringe passage to $1/50$ fringe or $\lambda/100$. For Hg green light ($\lambda = 5.46 \times 10^{-5}$ cm.) and spacers of length 1 cm. and expansion coefficient $\alpha = 25 \times 10^{-6}/\text{deg. C.}$ a temperature difference of only 2°C. would be required in order to measure the coefficient with an accuracy within 1%. This would enable the expansion coefficients to be measured to the same order of accuracy within a degree or two of the melting point or transition point, provided that the temperature of the spacers was suitably measured. The apparatus used in this investigation was merely a preliminary form of an accurate automatically recording dilatometer and was not capable of the above high order of accuracy.

The interferometric method, despite its obvious advantages, has been little used for expansion measurements on metal crystals. This is partly due to the necessity for continuous observation of the fringe movement over long periods of time, but, principally, because of the difficulty of construction of the three spacers. These must have the same expansion coefficient, and hence orientation, which, in practice, means that they must be cut from one single crystal.

A successful technique (Childs 1950) was developed for the preparation of such spacers from tin single crystals, and the distortion of the crystal lattice introduced during the process of construction of the spacers was kept sufficiently low to be of negligible effect on the thermal expansion.

§ 3. THE SINGLE-CRYSTAL SPACERS

(i) *Growth of the Tin Single Crystals*

The crystals were produced from 'Chempur' tin, in the form of rods 6 cm. long and 3 mm. diameter using a horizontal gradient furnace of the Kapitza (1928) type. This consisted of a rectangular copper bar of dimensions 30 cm. \times 4 cm. \times 0.7 cm., supported on asbestos knife edges in a draught-proof enclosure. An exponential temperature gradient of some $1.7\text{--}2.7^\circ\text{C./cm.}$ was maintained along its length by an electrical heater of resistance 77.5 ohm wound round one end of the bar. The rods of tin which were to be crystallized were laid on the surface of the bar with their axes in the direction of the maximum temperature gradient. Some of the rods were sealed into evacuated, and previously outgassed, glass tubes of slightly greater diameter than that of the rods. Others were placed directly on a clean Pyrex glass plate lying on the bar, the combined action of surface tension forces and the thick oxide coating serving to retain the shape of the rods even when molten.

The heater current was slowly reduced from a value at which the cool end of the rod was just molten to one at which the hot end became solid. This reduction was accomplished, using a manually operated rheostat, in steps of 0.005 amp. every 2–3 min., from 1.25 to 1.00 amp., corresponding to a minimum rate of growth of 1.8 mm/min. The thermal inertia of the bar was sufficient to cause a steady movement of the liquid–solid interface in the growing crystal. The percentage of crystals successfully grown in this way was rather low—35%—but was adequate for the investigation. No attempt was made, by seeding, to grow crystals of pre-determined orientations of 0 and 90°.

(ii) *Measurement of Orientation*

The orientations, ranging from 30 to 90°, of sixteen tin crystals were measured to within 1° by an optical method based on that of Chalmers (1936).

(iii) *Conversion into Spacers*

The distance of separation of the interferometer plates, chosen as a compromise between good definition of the fringes and large expansion of the crystal spacers, was 8.5 mm. Each crystal was cut into pieces of approximately this length, using a fine saw the set of which had been removed. Care was taken to avoid excessive pressure and impulsive stress during the cutting. The ends of the cut pieces were prevented from bending by a wooden backing strip to which the crystal had been glued.

After removal of the backing, the ends were sharpened separately using a fine flexible metal file, the pieces being held in the rubber-padded jaws of a lathe chuck. The lengths were measured with a travelling microscope and the pieces made equal in length to within 10^{-1} mm. by further filing.

At this stage the pieces of crystal, while substantially monocrystalline in form, were badly deformed at the ends. Much of the strained portion was then removed by etching each piece in concentrated hydrochloric acid until its length had been reduced by 0.2 mm. Finally, the lengths of all the pieces were made equal, to within 5 microns, by further periods of controlled etching and measuring.

By this technique crystal spacers were obtained with nearly all (more than 98%) of their volume apparently unaffected by the mechanical working. Some strain was still present, but once this had been released by annealing no differential expansion of the spacers, indicating changes in orientation, could be detected.

§ 4. THE INTERFEROMETER

The interferometer consisted of the two plates, P_1 and P_2 in Figure 1, separated by three spacers mounted in a holder. This holder was a brass ring 11 mm. internal diameter, and 2.5 mm. high. The spacers were located in grooves cut on the inside of the ring parallel to its axis, and were kept in position by individual leaf springs. Attached to the outside of the ring were three vertical guides of equal length. These served to locate the position of the upper plate without impedance to its motion caused by the expansion of the spacers.

The holder was inverted, and supported by the three guides on a flat glass surface. Each spring, in turn, was extended and a crystal spacer allowed to slide down the groove until it rested on the surface of a second piece of glass lying on the first. In this way the normal to the plane defined by the points of

the three spacers was made approximately parallel to the axis of the holder. The parallelism of the spacers was checked by using a coordinate travelling microscope and traversing, first along the length of a spacer, and then, at right angles along the holder. An angular deviation of a spacer from the axis of the holder of less than $\frac{1}{2}^\circ$ could thus be determined. After adjustment, the assembled spacer unit was stored in an inverted position until it was required for use.

The plates P_1 and P_2 were of optical glass 0.5 cm. thick. The upper plate P_1 was 1.3 cm. in diameter and its two optically flat surfaces were inclined at such an angle that light reflected at its top surface is thrown to one side. The lower plate P_2 , 3.8 cm. in diameter, was divided into two portions. One half supported the spacer unit and formed part of the expansion interferometer. The bottom

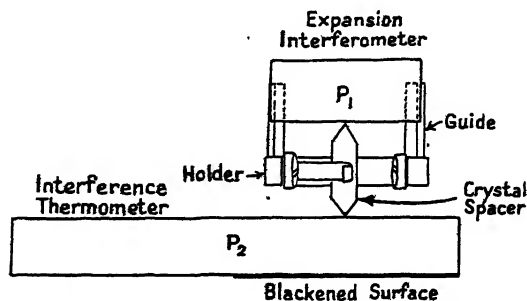


Figure 1. The assembled interferometer.

surface of this half was blackened to prevent reflection. The other half, the surfaces of which were inclined at an angle of 10 seconds, formed an interference thermometer of the type described by Luckeisch *et al.* (1922). This thermometer, though useful, was not of primary importance in the investigation.

The lateral movements of the fringes in the expansion interferometer and in the interference thermometer were observed relative to two sets of nylon cross hairs. These were attached under tension across the plates. Although not coinciding precisely with the location of the fringe systems no undesirable parallax was observed.

§ 5. APPARATUS

The dilatometer used for the measurement of the expansion of the single tin crystals is shown in Figure 2 (Plate). It consisted of a vacuum furnace A, shown diagrammatically in Figure 3, for heating the specimens and interferometer, and an autocollimating monochromator B, for producing a parallel beam of monochromatic light with which to illuminate the interferometer, and for viewing the passage of the fringe systems.

The inner furnace tube was made massive in order to ensure a slow, uniform rise in temperature for a given heater current, and the upper end was closed by an inclined circular plate-glass window which was sealed with Apiezon Q to the furnace tube and water-cooled. A rotary oil pump maintained a pressure of 10^{-2} to 10^{-3} mm. Hg.

The interferometer was mounted on a three-point support inside a thick copper bucket which, in its lowest position, rested on a metal cylinder standing on the base plate of the furnace tube. Through a central tube, one of two which protruded from the lower end of the furnace tube, there was a rod which served

to raise and lower the bucket, and a Wilson (1941) sliding seal at the end of this tube allowed free movement of the rod without the vacuum being broken. Through the other tube there was a copper-constantan thermocouple, the hot junction of which passed through a slot in the interferometer bucket and was at the same level as the crystal spacers. It was not possible to bring this junction into close contact with the spacers, and it was found, from a series of subsidiary tests, that the temperature of the spacers at 200°C . was approximately 3°C . higher than that recorded by the thermocouple. This is considered to be due to the upper interferometer plate acting as a shield against radiation loss. The thermocouple readings were accordingly corrected to give the true spacer temperatures.

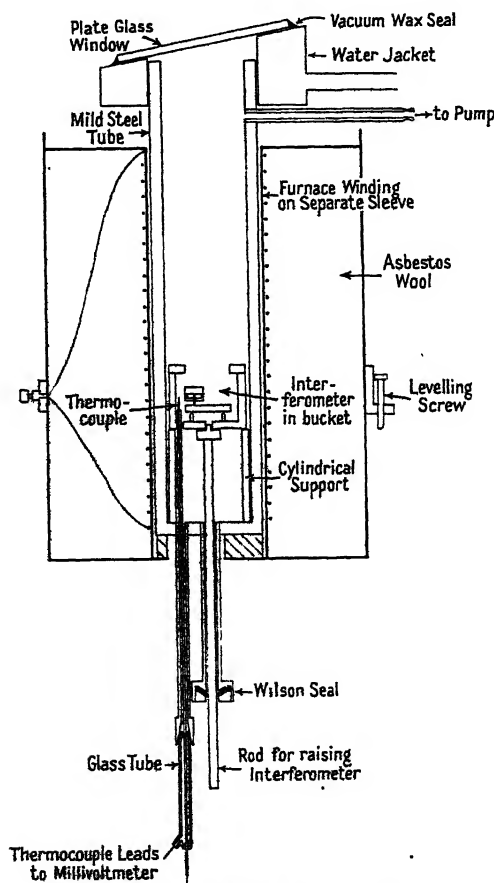


Figure 3. The vacuum furnace.

The thermo-E.M.F. was measured to 0.01 mv. (0.2°C .) with a millivoltmeter, the thermocouple being calibrated before, during and after each set of expansion measurements.

The furnace could be tilted without translation by three levelling screws mounted at the level of the interferometer.

The monochromator and viewing apparatus was a modified form of that described by Saunders (1945). The light from a mercury discharge tube which was situated perpendicular to, and slightly above, the horizontal axis of an

aplanatic lens combination, was reflected by a 45° front silvered mirror which formed a virtual image of the lamp just above the lens axis. A horizontal adjustable slit placed in the focal plane of the lens combination and in front of the lamp image limited the light and formed the true source, the length of which could be varied by means of a wedge diaphragm placed in front of it.

Emerging from the lens combination as a parallel beam, the light passed through a Pellin-Broca constant-deviation prism where it was dispersed and deviated downwards on to the horizontal plate surfaces of the interferometer. The paths of the reflected beams, which were slightly inclined to one another, were approximately coincident with the direction of the incident beam. These reflected beams passed through the prism, where they were further dispersed, and focused by the lens combination to pass through an aperture just below the slit source. Due to the dispersion of the prism, of the several overlapping images of the source formed at this point, only those of one particular wavelength (the mercury green $\lambda = 5460 \text{ \AA}$.) continued through to an adjustable eyepiece, where an image of the interferometer crossed by the two sets of interference fringes was formed (Figure 4, see Plate).

§ 6. EXPERIMENTAL PROCEDURE

(i) *Production of Fringes*

The following procedure was adopted in arranging the interferometer and spacers in position inside the furnace. The lower interferometer plate P_2 was placed on the three-point support in the copper bucket, and the bucket lowered by the rod to the bottom of the furnace. The bucket and plate were arranged so that the slot in the base of the bucket was vertically above the thermocouple tube, and the dividing line on the plate was parallel to the slit source. The furnace was then levelled by means of the levelling screws until the pair of images of the source formed by reflection at the plate appeared at the top of the aperture below the source. The field of view as seen through the eyepiece was bisected, the lower half being crossed by the horizontal temperature fringes.

The bucket was raised to the top of the furnace, and the spacer unit placed on the appropriate half of P_2 with the longest spacer nearest to the temperature fringes. The upper plate P_1 was then rested gently on the three spacers, with the cross-wires uppermost and at 45° to the dividing line of the lower plate. Great care had to be exercised to avoid moving the holder on the lower plate while the spacers were in contact with the glass since the points of the spacers were easily worn away by friction.

When the interferometer was lowered to the bottom of the furnace, two pairs of images of the source, one from the expansion interferometer and the other from the interference thermometer, were visible in the entrance aperture.

By slight adjustments of the level of the furnace these pairs of images were arranged so that one pair was at the top, and the two pairs symmetrically placed about the centre, of the aperture. Refraction at the wedge-shaped upper plate P_1 of the interferometer was the cause of the separation of the two pairs of images.

In general, fringes as in Figure 4 would then be visible across the interferometer. The spacing and orientation of these fringes naturally depended upon the precise differences in length of the three spacers. Slight uneven wear on their points, caused by vibration of the bucket when being lowered into position, frequently resulted in the number of fringes being too great for accurate observation of their

movement. To correct for this, first, the direction of the apex of the wedge-shaped film between the plates was found by identifying the images, due to reflection at the various surfaces, of a pin held over the interferometer. Slight vibration of the upper plate, produced by gentle tapping over the longest spacer with a pin attached to a length of thread, then sufficed to reduce the number of fringes to less than ten. After insertion of the thermocouple, the furnace window was waxed on, and the apparatus evacuated. Finally, in order to remove the last remnants of strain, the spacers were annealed above 200°C . for about five hours. Provided that the rotation and change in spacing of the fringes consequent upon the differential changes in length of the spacers during the release of strain were not such as to prevent accurate observation of the fringe passage relative to the cross-wires, the measurements of the expansion of the spacers could then be commenced.

(ii) *Method of Observation*

The movement of the interference fringes as the temperature of the furnace altered was observed visually and continuously, and the E.M.F. of the thermocouple corresponding to coincidence of each expansion fringe with the cross-wires was recorded. The alternative method of procedure, whereby the absolute length of the spacers at various steady temperatures is found using the method of fringe coincidences, could not be used since the intensities of the mercury lines, other than the green, emitted by the lamp were too low.

The rate of rise of temperature of the furnace was kept low in order to reduce the time lag in temperature between the spacers and the thermocouple. Further, by keeping the lag constant it disappeared from the calculation of the expansion coefficients. The rate used was 12°C/hr. , corresponding to a lag of about 1°C. , and this rate was maintained constant to within 20%, i.e. $\pm 0.2^{\circ}\text{C.}$ in the lag, by further small adjustments to the furnace current.

For each crystal the range of temperature from 20 – 225°C. was covered twice; generally, by one continuous run over the whole range, occupying some sixteen hours, and three shorter, overlapping, runs of 50 – 100°C. each. Observations were made for increasing temperatures only.

In calculating the expansion coefficients from the observational data the values of the thermocouple E.M.F. corresponding to coincidence of expansion fringe and cross-wires were found by interpolation from a smoothed E.M.F.–time curve.

The average expansion, at the cross-wires, of the three crystal spacers was next plotted as a function of temperature. The values of temperature corresponding to every tenth expansion fringe were read off from the smooth curve and the mean expansion coefficients over the intervals of 20, 30 and 40 fringes were calculated.

No direct use was made of the interference thermometer. It afforded much indirect practical assistance, however, since it ensured that if the thermocouple failed, an alternative temperature measuring device was available. Readings of the fractional order of the temperature fringes were taken at every coincidence of expansion fringe and cross-wires, and an independent calibration of the interference thermometer against the thermocouple was made.

Despite the careful annealing of the spacers before expansion measurements were taken, some slight further irreversible differential expansion of the spacers, amounting to one or two fringes over the whole range, often occurred. The

consequent rotation and change in spacing of the fringes was observed by noting periodically the number of fringes between pairs of the spacers.

At the completion of observations on each crystal, the spacers were re-measured, etched, and examined for signs of secondary growth.

§ 7. EXPERIMENTAL RESULTS

Six tin crystals of orientations $\theta = 86\frac{1}{2}^\circ$, $79\frac{1}{2}^\circ$, $44\frac{1}{2}^\circ$, $43\frac{1}{2}^\circ$, $31\frac{1}{2}^\circ$, and 30° were investigated. Except for the $79\frac{1}{2}^\circ$ crystal, the temperature interval $20\text{--}225^\circ\text{C}$. was covered twice; generally by one continuous run and three shorter overlapping runs. For the $79\frac{1}{2}^\circ$ crystal, the temperature interval $20\text{--}190^\circ\text{C}$. only was covered, in five overlapping ranges, readings being taken both for increasing and decreasing temperatures. The spacers in this case were unannealed and some rotation of the fringes occurred. As a check on the influence of this rotation on the thermal expansion, the observations over the lowest temperature range were repeated, and excellent agreement was found.

The expansion coefficients of the six crystals, as functions of temperature, are shown in Figure 5. On account of their large number, it is unfortunately

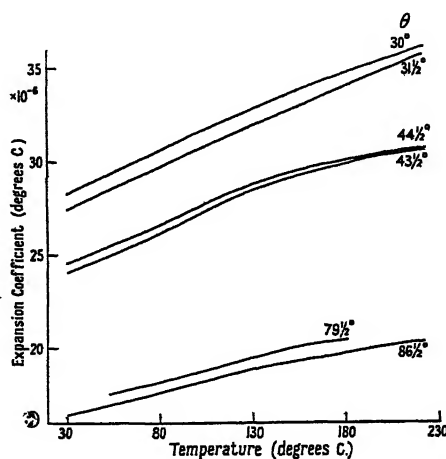


Figure 5. The variation of expansion coefficient with temperature for six tin crystals of different orientations.

impossible to show the individual experimental points, but only the smooth curves through them. The maximum deviation of a point from the corresponding curve is $1.2 \times 10^{-6}/\text{deg. c.}$ in the expansion coefficient, a 3% difference. The mean deviation is $0.3 \times 10^{-6}/\text{deg. c.}$

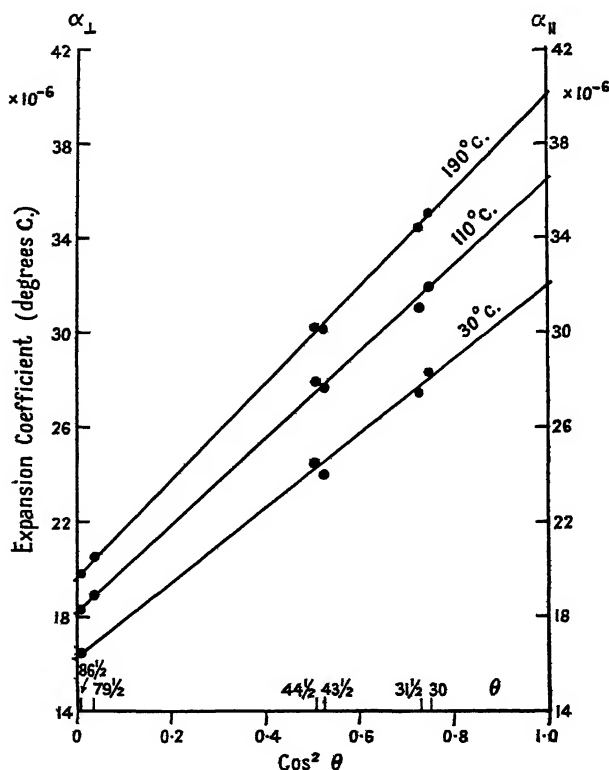
It will be seen from the figure that in each case the expansion coefficient increases by about 25% between room temperature and the melting point. In addition, the coefficients increase for decrease in θ , ranging at 30°C . from $16.4 \times 10^{-6}/\text{deg. c.}$ for $\theta = 86\frac{1}{2}^\circ$ to $28.3 \times 10^{-6}/\text{deg. c.}$ for $\theta = 30^\circ$.

The variation of the expansion coefficient with the cosine squared of the orientation is given, at 20°C . intervals, in Table 1, and shown graphically for three temperatures in Figure 6. This variation is linear, in agreement with Voigt's relation. Some idea of the accuracy of the observations may be obtained from the scatter of the experimental points which are completely independent. This scatter naturally includes errors in the orientation as well as in the expansion coefficients.

Table 1. The Expansion Coefficients of the six Tin Crystals at Various Temperatures

The values for $0^\circ (\alpha_{\parallel})$ and $90^\circ (\alpha_{\perp})$ have been obtained by extrapolation.

Temperature	Expansion coefficients ($\times 10^{-6}/\text{deg. c.}$) for tin crystals of given orientation							
($^\circ \text{C.}$)	0	30	$31\frac{1}{2}$	$43\frac{1}{2}$	$44\frac{1}{2}$	$79\frac{1}{2}$	$86\frac{1}{2}$	90
30	32.0	28.3	27.4	24.0	24.5	—	16.4	16.2
50	32.9	29.1	28.3	24.8	25.3	17.4	16.8	16.6
70	34.1	30.1	29.3	25.7	26.1	17.9	17.3	17.2
90	35.1	31.0	30.2	26.6	27.0	18.4	17.8	17.7
110	36.4	31.9	31.0	27.6	27.9	18.9	18.3	18.2
130	37.4	32.8	31.9	28.4	28.7	19.4	18.8	18.7
150	38.4	33.6	32.7	29.1	29.4	19.9	19.2	19.2
170	39.2	34.3	33.6	29.6	29.8	20.3	19.5	19.5
190	40.0	35.0	34.4	30.1	30.2	20.5	19.8	19.7
210	40.8	35.6	35.2	30.5	30.6	—	20.1	20.1
220	41.2	36.0	35.6	30.6	30.7	—	20.3	20.2

Figure 6. Variation of expansion coefficient with $\cos^2 \theta$ for three temperatures.

Finally, the values at 20°C. intervals of α_{\perp} and α_{\parallel} , obtained by extrapolation of curves such as those of Figure 6, are given in Table 1 and are shown as functions of temperature in Figure 7. Both coefficients increase with temperature, the increase being less marked at the higher temperatures. α_{\parallel} is almost exactly double α_{\perp} ; at 30°C. α_{\parallel} and α_{\perp} are 32.0 and $16.2 \times 10^{-6}/\text{deg. c.}$ respectively, and at 230°C. , by extrapolation, 41.5 and $20.3 \times 10^{-6}/\text{deg. c.}$ respectively.

The absence of any discontinuity in the curves of α_{\perp} and α_{\parallel} against temperature provides additional evidence against the existence of a rhombic form of tin which was once thought probable (Homer and Plummer 1939).

In Table 2 a comparison is made between the values for α_{\parallel} and α_{\perp} obtained in this investigation and the single values for the coefficients which have been

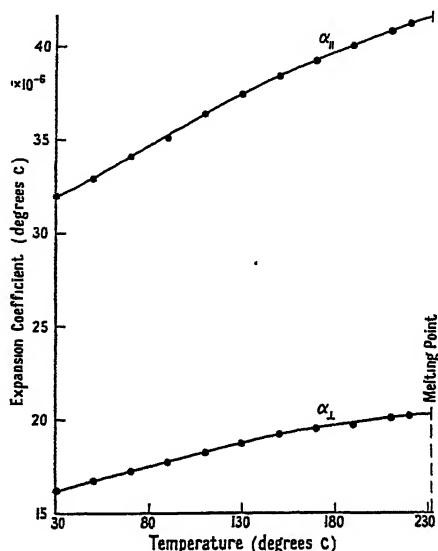


Figure 7. The principal expansion coefficients of tin as functions of temperature.

Table 2. Comparison between Values for α_{\parallel} and α_{\perp} for Tin found by various observers (O) and the authors (A).

Observer	Method	Temp. range (° C.)	Mean temp. (° C.)	$\alpha_{\perp} \times 10^{-6}/^{\circ}\text{C.}$		$\alpha_{\parallel} \times 10^{-6}/^{\circ}\text{C.}$	
				O	A	O	A
Bridgman (1925)	Optical lever	15-25	20	15.4	15.9	30.5	31.5
Kossolapow and Trapesnikow (1936)							
Trapesnikow (1936)	Debye-Scherrer x-ray	23-120	72	22.4	17.2	46.4	34.2
Shinoda (1933)	Debye-Scherrer x-ray	34-194	114	25.7	18.3	45.8	36.6

determined by other investigators. The discrepancies between the authors' values and those of Kossolapow and Trapesnikow (1936) and Shinoda (1933) are very much greater than the probable inaccuracy in the authors' values (see § 8). This may possibly be due to some effect arising from the Debye-Scherrer x-ray method which was used in both those investigations.

§ 8. ESTIMATION OF ACCURACY

The factors which contribute to the random experimental scatter present in the curves of Figure 6 can be divided into two groups: (i) those which involve errors of observation and calculation of the expansion coefficients of the spacers as assembled, and (ii) those which are due to the uncertainty in the crystallographic orientation corresponding to the measured coefficients of expansion.

Group (i). The total maximum error of determination of the temperature of the spacers at the moment of coincidence of an expansion fringe with the cross-wires was $\pm 0.6^\circ\text{C}$. This figure includes errors in the thermocouple calibration (0.2°C .), in the observation of the fringe coincidence and the reading of the millivoltmeter (0.2°C .), and due to the variation in the thermal lag of the spacers behind the thermocouple (0.2°C .). In practice, the combined error was rarely found to exceed $\pm 0.3^\circ\text{C}$. It was reduced to $\pm 0.2^\circ\text{C}$. or less by suitable choice of the (expansion fringe, temperature) curve. Hence, when temperature differences were taken to obtain the mean coefficient over a certain range of temperature a maximum error of $\pm 0.4^\circ\text{C}$. in the difference was possible. This corresponds, for a temperature interval of 40°C ., to an error in the observed expansion coefficient of $\pm 1\%$.

Group (ii). The orientations of the single crystals were measured to $\pm 1^\circ$. Further errors due to the residual strain and recrystallization remaining in the spacers after conversion from the crystal and, on assembly of the spacers in the interferometer, due to their departure from normality to the plate surfaces ($\frac{1}{2}^\circ$), also occurred.

If the measured orientation is assumed as correct, an additional maximum error of some $\pm 2\%$ must then be allowed for in the measured expansion coefficients of the individual crystals, thus making $\pm 3\%$ in all. By fitting the best line to the (expansion coefficient, $\cos^2\theta$) curves, the final error in α_\perp and α_\parallel was reduced to about $\pm 1\%$.

§9. CONCLUSIONS

Previous to this investigation the only reliable data for the variation of the coefficients α_\perp and α_\parallel with temperature for highly anisotropic metals were those for Zn and Cd. The variation of α_\perp and α_\parallel found for Sn is different from that exhibited by Zn and Cd, so that the data provide a further opportunity for the testing of the theoretical relationships such as that given by Grüneisen and Goens (1924).

ACKNOWLEDGMENTS

It is a pleasure to record our thanks to the College authorities, and to Professor A. M. Taylor, in whose department the work was carried out, for the facilities provided, also to the Department of Scientific and Industrial Research for the award of a maintenance grant to one of us (B. G. C.).

REFERENCES

- BRIDGMAN, P. W., 1925, *Proc. Amer. Acad. Arts Sci.* **60**, 305.
- CHALMERS, B., 1936, *Proc. Phys. Soc.*, **47**, 1.
- CHILDS, B. G., 1950, *J. Sci. Instrum.*, in the press.
- FIZEAU, A. H., 1869, *C.R. Acad. Sci., Paris*, **68**, 1125.
- GRÜNEISEN, E., and GOENS, E., 1924, *Z. Phys.*, **29**, 141.
- HOMER, C., and PLUMMER, H., 1939, *J. Inst. Met.*, **64**, 169.
- KAPITZA, P., 1928, *Proc. Roy. Soc. A*, **119**, 359.
- KOSSOLAPOW, G. F., and TRAPESNIKOW, A. K., 1936, *J. Exp. Theor. Phys. U.S.S.R.*, **6**, 577.
- LUCKEISCH, M., HOLLADAY, L., and SINDEN, R., 1922, *J. Franklin Inst.*, **194**, 251.
- SAUNDERS, J. B., 1945, *Bur. Stand. J. Res., Wash.*, **35**, 157.
- SHINODA, G., 1933, *Kyoto Imp. Univ. Proc. A*, **16**, 193.
- VOIGT, W., 1910, *Lehrbuch d. Kristallphysik*.
- WILSON, R. R., 1941, *Rev. Sci. Instrum.*, **12**, 91.

A Toroidal Magnetron*

By O. BUNEMAN

Atomic Energy Research Establishment, Harwell, Berks.

MS. received 24th August 1949

ABSTRACT. A rearrangement of the geometry of cavity magnetron oscillators is proposed such that the magnetic field loops around the cathode, whereas waves and electrons travel along it. Application of various magnetron theories to the new geometry leads to the result that the device will oscillate under conditions similar to those for a conventional cavity magnetron. The required magnetic field can be generated by a heavy current passed axially through the cathode.

§ 1. INTRODUCTION

A NEW kind of generator of microwaves is proposed here which resembles the conventional cavity magnetron in that it uses the action of crossed constant electric and magnetic fields upon a dense cloud of electrons for the purpose of exciting oscillations. The device differs from the conventional cavity magnetron chiefly in the orientation of the field and the oscillating cavities: a typical geometry is sketched in Figures 1(a) and (b).

There are advantages in this new type of geometry which may make the development of the device worth while. But it was not the outcome of a search for better microwave generators: it was conceived as the result of a general analysis of electron cloud behaviour under crossed fields. Possibly the only merit of the theory developed in connection with it will be a negative one, i.e. to serve as a warning as to what general oscillatory instabilities may arise in various electronic devices which employ crossed fields.

§ 2. PRINCIPLE OF THE TOROIDAL MAGNETRON

A sketch of the toroidal magnetron (the name 'torotron' might be proposed unless it has already been used elsewhere) is shown in Figures 1(a) and (b), but its principles can be more readily understood from the straightened-out version shown in Figure 2. The latter will be referred to as the 'axial flow magnetron'.

The cathode surface is cylindrical, as usual, and the anode is a concentric cylinder with resonator slots. The constant electric field, E , is radial, but the resonators, instead of being arranged azimuthally around the anode as in the conventional cavity magnetron, are arranged in a longitudinal succession as in the corrugated guide employed in linear accelerators.

Conversely, the magnetic field, H , instead of being longitudinal, loops around the cathode azimuthally. In fact the θ and z directions (Figure 2) have been interchanged relative to those in the conventional cavity magnetron.

The electrons, instead of circulating around the cathode, travel along it, at right angles to both electric and magnetic fields. The wave which is excited by them follows the same direction and is of the same type as that travelling in a linear accelerator. It will be shown that instability of the electron cloud to waves of a given frequency takes place under conditions very similar to those for conventional cavity magnetrons, i.e. similar voltages and magnetic field values

* Published by permission of the Controller, H.M. Stationery Office.

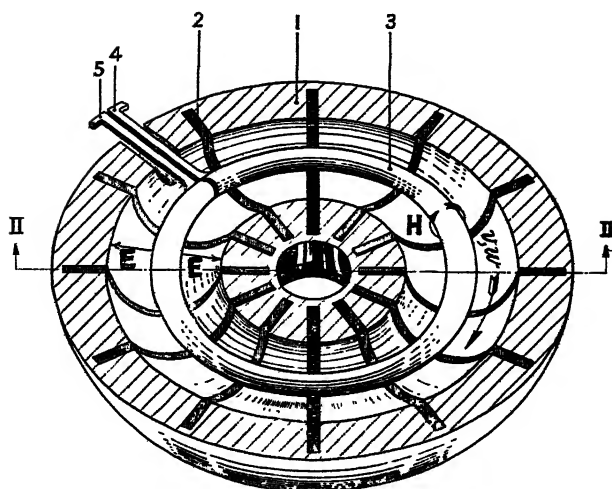
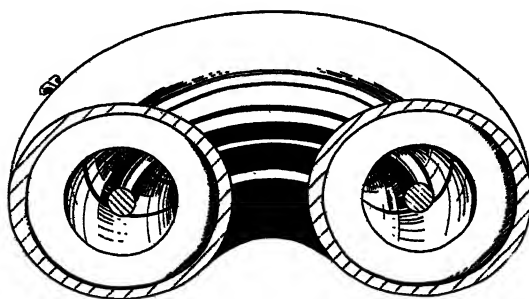


Figure 1 (a). Section of block and cathode.

- | | |
|----------------------|--------------------------|
| 1. Anode block. | E. Electric field. |
| 2. Resonator slots. | H. Magnetic field. |
| 3. Cathode. | v . Electron velocity. |
| 4, 5. Cathode leads. | w . Wave velocity. |



Section II-II

Figure 1 (b). Anode block showing resonators and cathode inside.

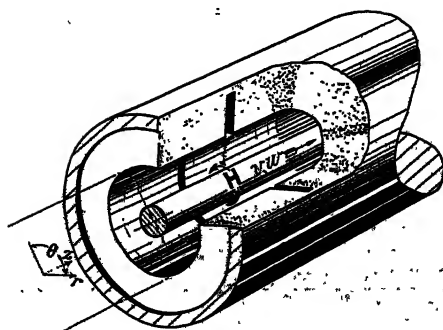


Figure 2. Axial flow magnetron.

must be employed: the interchange of the two directions in space only represents a minor distortion of the geometry. We shall derive approximate formulae for voltages, currents, efficiencies etc., and give reasons for hoping that very high power may be attained.

§ 3. ESTABLISHING THE MAGNETIC FIELD

A possible advantage of an azimuthal magnetic field is that it can be generated by a heavy current passed longitudinally through the cathode. The magnetic field due to such a current I_0 varies inversely as r :

$$H(r) = H_c/r = 2I_0/r \quad (\text{E.M.U.}) \quad \dots\dots(1)$$

where suffix c refers to values at the cathode. Using as units amperes, gauss and centimetres,

$$rH(r) = r_c H_c = I_0/5.$$

Typical values required would be: $H_c = 1,500$ gauss, $r_c = 0.5$ cm. Hence $I_0 = 3,750$ amp., current density $= 15,000/\pi$ amp/cm², giving a field of 0.01 volt/cm. along a water-cooled copper conductor placed inside the cathode ($\sigma = 450,000$ mho/cm.). The heat developed is therefore 40 watts per cm. length, which is easily removed by water-cooling. For other values of H_c (gauss), r_c (cm.) and σ (mho/cm.) one should use the formulae: current density (amp/cm²) $= 5H_c/\pi r_c$, field (volt/cm.) $= 5H_c/\pi r_c \sigma$, heating (watt/cm.) $= 25H_c^2 \pi \sigma$.

In pulsed operation the full field is required only during a fraction of the time, and the dissipation problem might disappear. The current pulses required to establish the field might then be used (and adjusted in length) to supply normal cathode heating.

§ 4. THE STATIC MAGNETRON

(i) Cut-off

We hope to show that the new device is likely to operate, and we wish to find the approximate operating conditions. We therefore modify some of the theories which were developed in connection with the ordinary magnetron (Collins 1948) so that they apply to the new geometry (straight version). We begin with the static case.

A cut-off potential can be derived for the condition when the electric field is static and purely radial; the equation for the axial acceleration is then

$$\frac{d^2 z}{dt^2} = \frac{dr}{dt} \frac{eH}{mc} = \frac{eH_c r_c}{mc} \frac{dr}{dt} \frac{1}{r},$$

giving, by integration,
$$\frac{dz}{dt} = \frac{eH_c r_c}{mc} \ln \frac{r}{r_c} \quad \dots\dots(2)$$

for zero emission velocity at the cathode. The potential relative to the cathode must therefore exceed a value $V(r)$ given by

$$\frac{eV(r)}{mc^2} = \frac{1}{2} \left(\frac{eH_c r_c}{mc^2} \right)^2 \left(\ln \frac{r}{r_c} \right)^2. \quad \dots\dots(3)$$

This takes the place of Hull's (1921) cut-off formula, into which it transforms on introducing the approximation

$$\ln \frac{r}{r_c} \simeq \frac{1}{2} \left(\frac{r}{r_c} - \frac{r_c}{r} \right), \quad \dots\dots(4)$$

correct to cubes of $\ln(r/r_c)$.

In a (V, H_c) plot one obtains a parabola. Likewise a $(V, \ln r)$ plot for constant H_c yields a parabola (as against a catenary for ordinary magnetrons). In practical units we use

$$mc^2/e = 510 \text{ kv.} = 1,700 \text{ gauss.cm.,}$$

and for a typical example we choose H_c, r_c as in §3, and $r = 1.36 \text{ cm.}$, so that $\ln(r/r_c) = 1.00$. We then obtain $V = 50 \text{ kv.}$ for cut-off.

(ii) *Single Stream Steady State*

When the anode voltage is below cut-off, a self-consistent, zero current steady state can be obtained in which all electrons move axially only ($dr/dt = 0, d\theta/dt = 0$) with velocity dz/dt given by (2), so that the potential follows the $(V, \ln r)$ parabola given by (3) as far as the stream extends, i.e. from the cathode, $r = r_c$, to the electron cloud surface, $r = r_s$. Beyond this there is no charge and the potential varies linearly with $\ln r$. Continuity of the field at $r = r_s$ requires the potential line to be tangential to the parabola there. Thus, writing l for $\ln(r/r_c)$,

$$V = V_s + \left(\frac{dV}{dl} \right)_s (l - l_s)$$

or

$$\frac{eV}{mc^2} = \frac{1}{2} \left(\frac{eH_c r_c}{mc^2} \right)^2 [l_s^2 + 2l_s(l - l_s)] = \frac{1}{2} \left(\frac{eH_c r_c}{mc^2} \right)^2 [l^2 - (l - l_s)^2] \quad \dots\dots (5)$$

beyond the cloud. For given anode voltage less than cut-off, l_s , and hence the cloud radius, can be found by constructing on the (V, l) plot the tangent to the cut-off parabola from the point (V_a, l_a) representing the anode, or by solving (5) for l_s .

Assuming H_c, r_c and $r = r_a$, as in the example above, but an anode voltage of, say, 42 kv., we obtain $l_s = 0.60, r_s = 0.91 \text{ cm.}$

The confined electron beam has the following further properties:

Electronic density:

$$\rho = \frac{1}{4\pi} \nabla^2 V = \frac{mc^2}{4\pi e} \left(\frac{eH_c r_c}{mc^2} \right)^2 \frac{1}{r^2}. \quad \dots\dots (6)$$

Current density:

$$j = \rho \frac{dz}{dt} = \frac{mc^3}{4\pi e} \left(\frac{eH_c r_c}{mc^2} \right)^3 \frac{\ln(r/r_c)}{r^2}.$$

Beam current:

$$I = \int_{r_c}^{r_s} 2\pi r j dr = \frac{mc^3}{4\pi e} \left(\frac{eH_c r_c}{mc^2} \right)^3 \pi \left(\ln \frac{r_s}{r_c} \right)^2. \quad \dots\dots (7)$$

To obtain the beam current in practical units we use

$$mc^3/4\pi e = 1,350 \text{ amp.,}$$

and we get 130 amp. in the quoted example. In some of the resonance considerations we require also the 'plasma electron frequency', which is $(ep/\pi m)^{1/2}$ (see, for instance, Stratton 1941, ch. v, 5.16, equation 27). By substitution from (5) we obtain

$$\nu_{\text{plasma}} = \frac{1}{2\pi} \frac{eH_c r_c}{mcr}, \quad \dots\dots (8)$$

i.e. the 'cyclotron frequency' at radius r and the plasma frequency are identical.

The axial beam described here is the counterpart to Brillouin's (1941) circulating steady state in ordinary magnetrons. Most of the formulae quoted are the same as for Brillouin's state, except for correction factors which differ from unity by quantities of the order $[\ln(r/r_c)]^2$. There is an essential difference in the total beam current, in that length and circumference are interchanged in the two types of magnetron: the beam current, in Brillouin's state, is roughly $(L/2\pi r)$ times that given above, L being the axial length.

§ 5. THE OPERATING MAGNETRON

(i) Threshold Voltage

The threshold voltage which the author derived for conventional magnetrons (Manchester Group 1942, Collins 1948, pp. 30, 31, 238, 340, 341) can be obtained more easily in the present case. One assumes the electric fields to be neither purely radial nor static, but of the travelling wave type, so that they appear static (yet still not purely radial) when viewed from a frame of reference travelling with the wave.

In such a frame the equations of motion are the same as if the frame were stationary, except for a force which arises from the fact that all electrons, in addition to their velocity relative to the frame, share the frame velocity w . This force is $ewH(r)/c = ewH_c r_c/cr$, and is directed inwards when the wave travels in the same direction as the main electron flow. It has a potential $e(w/c)H_c r_c \ln(r/r_c)$. We must therefore subtract $(w/c)H_c r_c \ln(r/r_c)$ from the electrostatic potential V and can then write the equations of motion as if the frame were stationary. Particles are emitted with kinetic energy $\frac{1}{2}mw^2$ in the travelling frame, and any change in their kinetic energy must be due to the electrostatic potential, modified as described. In all regions reached by particles we must have, therefore,

$$eV - \frac{w}{c} eH_c r_c \ln \frac{r}{r_c} + \frac{1}{2}mw^2 \geq 0.$$

The threshold voltage V_T is the lowest anode voltage for which this inequality is satisfied:

$$\frac{eV_T}{mc^2} = \frac{w}{c} \frac{eH_c r_c}{mc^2} \ln \frac{r_a}{r_c} - \frac{1}{2} \left(\frac{w}{c} \right)^2 \quad \dots\dots (9)$$

or

$$\frac{V_T}{510 \text{ kv.}} = \frac{\lambda^*}{\lambda} \frac{H_c r_c}{1,700 \text{ gauss.cm.}} \ln \frac{r_a}{r_c} - \frac{1}{2} \left(\frac{\lambda^*}{\lambda} \right)^2, \quad \dots\dots (9')$$

where λ^* is the wavelength in the guide, λ that in free space, so that $\lambda^*/\lambda = w/c$. In a typical case we might choose $\lambda = 10$ cm., $\lambda^* = 2$ cm., H_c , r_c and r_a as above and obtain $V_T = 35$ kv. At this voltage the steady state surface (see equation (4)) lies at $l = 0.45$ or $r = 0.784$ cm.

Equations (9) and (9') represent straight lines on a (V_T, H_c) diagram, tangential to the cut-off parabola. The approximation (4) changes (9) into the threshold formula for conventional magnetrons, provided one remembers that there the guide wavelength is $2\pi r_a$ divided by n , the number of repeats of the wave around the block.

(ii) Synchronism

Equation (9) may be rewritten

$$\frac{eV_T}{mc^2} = \frac{1}{2} \left(\frac{eH_c r_c}{mc^2} \right)^2 \left[l_a^2 - \left(l_a - w \sqrt{\frac{eH_c r_c}{mc}} \right)^2 \right],$$

which, on comparison with (5), shows that when $V_a = V_T$, $w = \frac{eH_c r_c}{mc} l_a =$ axial velocity of the surface stream of the steady beam (see equation (2)). In the axial flow magnetron the threshold condition and the condition of synchronism between wave and outer electrons are identical. In the conventional magnetron this is *not* so: under threshold conditions the outer electrons are faster than the wave (see the author's Manchester report 1942).

It is reasonable to expect that, in order to do work on the radio-frequency field, the electrons have to be faster than the wave, i.e. to push the wave along. The threshold formula may therefore underestimate the voltage necessary for oscillation, and we proceed to explore alternative voltage criteria.

(iii) 'Plasma Resonance'

A resonance phenomenon which was discussed by the author in another Manchester report (Manchester Group 1943) arises from the interaction between electrons and waves which occurs when the electron stream encounters the wave crests at its plasma frequency (see also Collins 1948, p. 260). We assume the electrons to be overtaking the waves so that the velocity of encounter is $(eH_c r_c l / mc) - w$. The wavelength in the guide is λ^* and the condition for plasma resonance is

$$\frac{eH_c r_c l}{mc} - w = \lambda^* \frac{1}{2\pi} \frac{eH_c r_c}{mcr}. \quad \dots\dots (10)$$

With some algebraic manipulation this can be reduced to

$$\frac{2\pi r}{\lambda^*} \exp(-\lambda^*/2\pi r) = \frac{2\pi r_0}{\lambda^*}, \quad \dots\dots (11)$$

where r_0 = radius of synchronous layer, r = radius of resonant layer. Given λ^* , λ and $H_c r_c$, one can obtain r_0 from (2) and then r from (11) or from the following table:

Radius of Layer in Plasma Resonance, r , and Radius of Synchronous Layer, r_0 ,
(both in units of $\lambda^*/2\pi$)

$2\pi r_0/\lambda^*$	0	0.5	1.0	1.5	2.0	2.5	3.0	3.5	4.0	4.5	5.0	large
$2\pi r/\lambda^*$	0	1.172	1.764	2.311	2.842	3.365	3.881	4.394	4.904	5.413	5.920	$\frac{2\pi r_0}{\lambda^*} + 1$

The new value of r can then be fed into (5) to give the anode voltage at which the electron cloud is wide enough for the surface to be in plasma resonance. In our particular case we have seen that the synchronous radius is 0.784 cm., and as $\lambda^* = 2$ cm., we get $\ln(r/r_0) = 0.75$, which, from formula (5), corresponds to an anode voltage of 47 kv.—very near cut-off and well above threshold.

(iv) *Small Amplitude Theory*

The small amplitude theory (Collins 1948, §6.7, Manchester Group 1944) can be applied to the axial flow magnetron. It has not been followed through in detail, but an outline of the procedure is given in the Appendix. The entire analysis is very similar in the two types of magnetron and the change of geometry results only in minor modifications in some of the perturbation equations. We can therefore predict with confidence that the new type of valve will oscillate and that spontaneous response (exponential growth of amplitudes from the background of thermal noise) will occur when the electrons are 'matched' to the load, somewhere between the conditions of synchronism and 'plasma resonance'.

Exact matching requires knowledge of the Q of the load. To avoid lengthy calculations we choose for the radius of a 'matched' electron cloud in our particular example the value $r_s = 0.91$ giving $l_s = 0.60$, which is half-way between the synchronous and resonant values of l (0.45 and 0.75). It corresponds to 42 kv. anode voltage (see §4 (ii)).

In the ordinary cavity magnetron, threshold and charge cloud instability due to matching occur at very close levels, and one cannot easily decide which of the two is relevant. Here, the threshold level, being identical with synchronism, is always below the instability level. Thus the crests of the spontaneous charge cloud waves will always be able to penetrate to the anode. It is possible that the success of the threshold condition was only fortuitous, owing to its proximity to the (rather more complicated) instability condition. It is unlikely that oscillation will begin without instability. If the threshold condition is only a necessary but not a sufficient condition we shall be able to disregard it in the new magnetron. On the other hand, there may be some deeper significance in the proximity between threshold and instability in ordinary cavity magnetrons (perhaps in connection with moding problems). In that case the new device would fail.

§6. ELECTRONIC CURRENTS

Theory is, unfortunately, unable to make reliable predictions of the amount of electronic current which the magnetron will pass under operating conditions. The nearest approach to calculating currents in ordinary magnetrons was achieved in the self-consistent field and orbit calculations done by the Manchester team under Hartree's direction and in similar calculations carried out at Leeds under Stoner. These required an immense amount of numerical work in each particular case.

A rough estimate—probably an upper limit—of the current can be obtained from the argument that under oscillatory conditions of large amplitude something like the axial beam current of §4(ii) will be diverted to the anode in every wavelength λ^* . In our example this would lead to 130 amp. per wavelength, the total for a valve of axial length L being $130 L/\lambda^*$ amp. The question is how long (in terms of number of wavelengths) the magnetron can be made.

In an ordinary magnetron of similar dimensions we should have obtained, by the same argument, total current = circulating current in Brillouin's state times number of waves around the block = $130 (L/2\pi r)n = 130 (L/2\pi r)(2\pi r/\lambda^*) = 130 L/\lambda^*$, as above.

One restriction on length is absent in the new device: that arising from space limitations of the magnetic field, and we might thus gain a real advantage in the new type of valve.

§ 7. EFFICIENCIES

Formulae for an ideal efficiency were given by the author (Manchester Group 1942) for conventional magnetrons. Similar formulae apply in the present case: particles which reach the anode at threshold condition ($V_a = V_T$) are stationary with respect to the wave, i.e. travel with velocity w relative to fixed axes. Of the total energy eV_T which they have acquired from the D.C. field they have retained $\frac{1}{2}mw^2$ and surrendered the balance, $eV_T - \frac{1}{2}mw^2$, to the radio-frequency field. At operation appreciably above threshold ($V_a > V_T$) one would guess that the electrons carry away the excess $eV_a - eV_T$, still leaving only $eV_T - \frac{1}{2}mw^2$ for the radio-frequency field. The efficiency is thus

$$\eta \simeq \frac{eV_T - \frac{1}{2}mw^2}{eV_a}.$$

For our example ($V_T = 35$ kv., $V_a = 42$ kv., $w = 0.2c$) we obtain 58%.

§ 8. OUTPUT POWER

The input power in our example is 2.7 MW. per cm. length: hence an output of perhaps up to $1\frac{1}{2}$ MW. per cm. might be anticipated, thus if we assume five internal wavelengths along the magnetron ($L = 5\lambda^* = 10$ cm.), we obtain 15 MW. The procedure for calculating outputs for different dimensions, fields and frequencies is obvious. If the ordinary magnetron of similar anode and cathode radii could be made as long as 10 cm. the output would be similar.

§ 9. CONTINUOUS ELECTRON BEAM IN TOROIDAL GEOMETRY

The reason for proposing a toroidal arrangement rather than a straight arrangement is the necessity of establishing continuity of the electron stream. The electrons drift along equipotentials in crossed fields. If an equipotential is interrupted by an insulator (as it would have to be in the straight device of finite length), the insulator may be destroyed unless elaborate 'dumping' arrangements are made for the electrons. In the toroidal device the leads could be shaped for partial or complete mutual annulment of their magnetic fields in the vicinity of the seals through which they are brought out. The seals could thus be protected from bombardment, and perhaps further by not coating the cathode 'upstream' of the roots of the leads. An alternative to the toroidal arrangement would be the 'race track', i.e. two straight sections joined by semicircular electron- and waveguides.

§ 10. MODE PROBLEMS

A certain limitation to the length, and therefore to the output, will be set by the condition of mode stability. Separation of the wave velocities, i.e. of λ^*/λ , is of importance in the threshold and synchronism conditions. But in the plasma resonance formula λ^* and λ figure individually, and it is not possible to say that high dispersion of wave velocities only is of importance.

The spectrum of modes in the projected valve has not yet been studied. In the first place one ought to investigate the case of a straight corrugated guide of finite length with a periodicity condition at the ends, in approximation to the corrugated torus. If azimuthal modes in the guide can be ignored, there is hope that the length might be made quite substantial, i.e. many guide wavelengths long, without excessive closeness in the wave velocities of the various modes. There should exist a considerable amount of experience in this field, from studies of the linear accelerator. The principle of the 'rising sun' can, of course, be applied. There is, however, no obvious way of translating the idea of strapping.

§ 11. CONCLUSIONS

Having shown that the new type of magnetron geometry is equivalent to the old type as far as charge cloud instability and small amplitude phenomena are concerned, we may be confident that the device will oscillate. We have seen also that most of the more elementary calculations on magnetron space-charge behaviour can be translated with ease. Conditions of operation can then be derived which are in most respects similar to the usual magnetron conditions. There is one qualitative difference which may turn out to be significant, in that the threshold voltage is identical with that for which the electron cloud surface is only just as fast as the wave, whereas charge-cloud instability theory requires the electrons to be somewhat faster. However, there is no known reason why this should prevent operation at instability conditions.

The power derivable from the new device depends on the length of the cathode which, in turn, depends on the stability against unwanted modes. It is hoped that the length can be increased beyond that in the ordinary magnetron.

ACKNOWLEDGMENT

Acknowledgment is made to the Director of the Atomic Energy Research Establishment for permitting publication of this paper.

REFERENCES

- BRILLOUIN, L., 1941, *Phys. Rev.*, **60**, 385.
 COLLINS, G. B., 1948, *Microwave Magnetrons* (New York: McGraw-Hill), § 6.
 HULL, A. W., 1921, *Phys. Rev.*, **18**, 31.
 LAMB, H., 1945, *Hydrodynamics* (New York: Dover Publications), 6th edition.
 MANCHESTER GROUP, 1942, *C.V.D. Mag.*, **10**, **11**, **17**; 1943, *Ibid.*, **31**; 1944, *Ibid.*, **37**.
 STRATTON, J. A., 1941, *Electromagnetic Theory* (New York: McGraw-Hill).

APPENDIX

SMALL AMPLITUDE THEORY

The instabilities of the Brillouin charge cloud were obtained by the author, using a perturbation method (Collins 1948, § 6.7). Unfortunately, the simple mechanical analogy of a rotating platform which is used in the Manchester Group report (1944) in order to simulate the action of the magnetic field cannot be taken over into the theory of the axial flow magnetron as the magnetic field is no longer uniform. A more complicated analysis is necessary before a velocity potential can be introduced and before the hydrodynamical approach of the above report can be repeated.

The equation of motion is

$$m \frac{d\mathbf{v}}{dt} = e \left(\text{grad } V - \frac{\mathbf{v}}{c} \wedge \text{curl } \mathbf{A} \right), \quad \dots\dots (A1)$$

where \mathbf{A} is the vector potential of the magnetic field. \mathbf{A} has only an axial component A_z of value $H_0 r_0 \ln(r/r_0)$. We apply Eulerian methods to express $m d\mathbf{v}/dt$ in terms of partial derivatives:

$$\begin{aligned} m \frac{d\mathbf{v}}{dt} &= m \frac{\partial \mathbf{v}}{\partial t} + m(\mathbf{v} \cdot \text{grad}) \mathbf{v} \\ &= m \frac{\partial \mathbf{v}}{\partial t} + \text{grad} \left(\frac{1}{2} m v^2 \right) - \mathbf{v} \wedge \text{curl } m\mathbf{v}. \quad \dots\dots (A2) \end{aligned}$$

We introduce the generalized momentum for electrons in a magnetic field,

$$\mathbf{p} = m\mathbf{v} - e\mathbf{A}/c, \quad \dots\dots(A3)$$

and obtain, by combining (A1) and (A2), the following equation for \mathbf{p} :

$$\frac{\partial \mathbf{p}}{\partial t} = \mathbf{v} \wedge \text{curl } \mathbf{p} + \text{grad} (eV - \frac{1}{2}mv^2). \quad \dots\dots(A4)$$

In the ordinary magnetron \mathbf{p} is simply the momentum relative to the Larmor frame of reference, where we know \mathbf{p} to be irrotational. We try to prove that \mathbf{p} is irrotational in the axial flow magnetron by studying $\text{curl } \mathbf{p}$, which we denote by \mathbf{q} . We can find an equation for the change of \mathbf{q} with time, following the electrons:

$$\begin{aligned} \frac{d\mathbf{q}}{dt} &= \frac{\partial \mathbf{q}}{\partial t} + (\mathbf{v} \cdot \text{grad}) \mathbf{q} = \text{curl} \left(\frac{\partial \mathbf{p}}{\partial t} \right) + (\mathbf{v} \cdot \text{grad}) \mathbf{q} \\ &= \text{curl} (\mathbf{v} \wedge \mathbf{q}) + (\mathbf{v} \cdot \text{grad}) \mathbf{q} \quad (\text{from (A4)}) \\ &= (\mathbf{q} \cdot \text{grad}) \mathbf{v} - \mathbf{q} \text{div } \mathbf{v}. \end{aligned}$$

This equation shows that the property of zero vorticity, provided it occurs at all, is retained and carried along by the electrons. As no vortices are being emitted from the cathode (the proof is similar to that for ordinary magnetrons), no vortices will be formed and $\text{curl } \mathbf{p}$ will vanish everywhere.

The vector \mathbf{p} can therefore be expressed as the gradient of a scalar to be denoted mF . (The quantity F is then equivalent to that in the Manchester Group report (1944), where it appears as the velocity potential in the Larmor Frame.)

By substituting $\mathbf{p} = \text{grad } mF$ in (A4), and integrating, one obtains the generalized Bernoulli equation (Lamb 1945, Ch. II, Art. 17):

$$\frac{2eV}{m} = v^2 + 2 \frac{\partial F}{\partial t}. \quad \dots\dots(A5)$$

The steady state axial beam is characterized by $F=0$ and hence $\mathbf{p}=0$. We study undulatory perturbations of this state, i.e. perturbations whose z and t variation is given by the factor $\exp \{2\pi i(\omega t - z)/\lambda^*\}$, so that $\partial/\partial t = 2\pi i\omega/\lambda^*$ and $\partial/\partial z = -2\pi i/\lambda^*$. We put

$$F = \delta F = f(r) \exp \{2\pi i(\omega t - z)/\lambda^*\}$$

and derive for the perturbations of velocities, potentials and densities (unperturbed values are given in §4(ii)):

$$\delta v_r = v_r = df/dr,$$

$$\delta v_z = 2\pi if/\lambda^*,$$

$$\delta V = \frac{iH_0}{c} \left(\frac{2\pi mc^2}{eH_0\lambda} - \frac{2\pi r_0}{\lambda^*} \ln \frac{r}{r_0} \right) f,$$

$$\delta \rho = -\frac{1}{4\pi} \left(\frac{1}{r} \frac{d}{dr} \left(r \frac{d}{dr} \right) - \frac{4\pi^2}{\lambda^{*2}} \right) \frac{iH_0}{c} \left(\frac{2\pi mc^2}{eH_0\lambda} - \frac{2\pi r_0}{\lambda^*} \ln \frac{r}{r_0} \right) f,$$

omitting in each case the exponential factor,

The equation of continuity (first-order perturbations) becomes

$$\frac{1}{r} \frac{\partial}{\partial r} (r \rho_{\text{unp}} \delta v_r) + \frac{\partial}{\partial z} \rho_{\text{unp}} \delta v_z + \frac{\partial}{\partial z} v_{z, \text{unp}} \delta \rho + \frac{\partial \delta \rho}{\partial t} = 0,$$

which, on substitution, yields a second-order differential equation for f :

$$r \frac{d}{dr} \frac{r_c^2}{r^2} \frac{df}{dr} - \frac{4\pi r_c^2}{\lambda^{*2}} f - \left(\frac{2\pi m c^2}{e H_c \lambda} - \frac{2\pi r_c}{\lambda^*} l \right) \left(\left(r \frac{d}{dr} \right)^2 - \frac{4\pi r^2}{\lambda^{*2}} \right) \left(\frac{2\pi m c^2}{e H_c \lambda} - \frac{2\pi r_c}{\lambda^*} l \right) f = 0.$$

This can be simplified to

$$\frac{d}{dl} g(l) \frac{df}{dl} = h(l) f \quad \dots\dots (A 6)$$

$$\text{where } g(l) = e^{-2l} - \left(\frac{2\pi m c^2}{e H_c \lambda} - \frac{2\pi r_c}{\lambda^*} l \right)^2 \quad \text{and} \quad h(l) = \left(\frac{2\pi r}{\lambda^*} \right)^2 g(l).$$

This equation is of the same form as the perturbation equation for ordinary magnetrons. To change to the latter from (A 6) one must replace λ^* by $2\pi r/n$ and multiply some of the terms in g and h by factors of the type $1 + O(l)^2$ or $1 \pm 2/n^2$. These are minor modifications which will not change the qualitative features of the equation. Of these, the most important are the zeros of $g(l)$, i.e. the singularities which occur at the plasma resonances (see § 5, (iii)) and at which the general solution becomes logarithmically infinite, as in the ordinary magnetron.

Calculation of the charge-cloud impedance from the perturbation theory will thus give results as in the Manchester Group report (1944): negative resistance beyond the singularity, reactance varying inversely with frequency before the singularity is reached, so that the resistance in the load can be matched to the electrons by assuming an exponential increase of amplitudes (Manchester Group 1944, Collins 1948, § 6.7).

Viscous Flow Transverse to a Circular Cylinder

By C. N. DAVIES

(Member of Scientific Staff, Medical Research Council),
London School of Hygiene and Tropical Medicine, Keppel Street, London

MS. received 12th September 1949

ABSTRACT. Expressions for the flow field about a circular cylinder are derived which are valid for small values of the Reynolds' number. These are obtained by Lamb's method from Oseen's equation, and it is shown that the formulae previously given by Lamb involve the neglect of a first order term. Lamb's expression for the drag is correct to the first order in the Reynolds' number.

IF the Navier-Stokes equations of hydrodynamics are linearized by neglecting the inertia terms it is impossible to obtain a steady state solution with finite velocity at infinity for the two dimensional flow of incompressible fluid past a circular cylinder. By linearizing according to Oseen's method, however, Lamb (1911) obtained an approximate solution. Later, Bairstow, Cave and Lang (1923) and Faxén (1927) published exact solutions. These, however, are complicated and since the formulation of Oseen's linearized equation involves

assumptions which are only acceptable, in general, at low values of the Reynolds' number, R , a simple approximate formula, accurate to the first order in R , is all that is really required.

When Lamb's approximation was examined in connection with the application of his results to the theory of fibrous air filters it was found to be incorrect owing to the unjustifiable neglect of a term of the first order in R . In the present paper Lamb's method is followed and expressions for the field of flow accurate to the first order in R are derived. It is shown that his well-known formula for the drag on a cylinder is correct.

The general dynamical equation for an incompressible fluid in the absence of body forces is

$$\frac{\partial \mathbf{Q}}{\partial t} + (\mathbf{Q} \cdot \nabla) \mathbf{Q} = -\frac{1}{\rho} \nabla p + \nu \nabla^2 \mathbf{Q}, \quad \dots\dots(1)$$

where \mathbf{Q} and p , the velocity and pressure in the fluid, are functions of position, ρ is the density and ν is the kinematic viscosity. For steady flow with complete neglect of the inertia term the right-hand side is equal to zero. As an alternative, Oseen replaced \mathbf{Q} by $U\mathbf{i} + \mathbf{q}$, where $U\mathbf{i}$ represents the velocity of the undisturbed stream and \mathbf{q} the perturbation produced by an obstacle. Now only non-linear terms in \mathbf{q} are neglected, so that for steady flow equation (1) reduces to

$$(U\mathbf{i} \cdot \nabla) \mathbf{q} = U \frac{\partial \mathbf{q}}{\partial x} = -\frac{1}{\rho} \nabla p + \nu \nabla^2 \mathbf{q}. \quad \dots\dots(2)$$

It can be shown (Lamb 1932) that this leads to a more accurate representation of the flow field at a distance from the obstacle than is obtained if the inertia term $(\mathbf{Q} \cdot \nabla) \mathbf{Q}$ is completely neglected. The effect of the approximation introduced into the fundamental equation can be regarded as the change produced by a small body force, acting upon the fluid, which is a function of position. Complete neglect of the inertia term means that only viscous forces are assumed to act on each element of fluid; in this case the compensating body force tends to zero at the surface of the obstacle so that the approximation is a good one near the surface. Oseen's partial allowance for inertia leads to the result that the body force does not vanish at the surface, but its ratio to the viscous force is of the same order of magnitude as the Reynolds' number R . Hence, if R is small, the result is satisfactory near the obstacle as well as at a distance from it.

Putting $k = U/2\nu$ equation (1) becomes

$$\frac{1}{\rho} \nabla p = \nu \left(\nabla^2 - 2k \frac{\partial}{\partial x} \right) \mathbf{q}. \quad \dots\dots(3)$$

We have also the equation of continuity

$$\nabla \cdot \mathbf{q} = 0. \quad \dots\dots(4)$$

Differentiating (3) gives $\nabla^2 p = 0$ of which a solution is

$$p = \rho U \partial \phi / \partial x, \quad \dots\dots(5)$$

provided

$$\nabla^2 \phi = 0. \quad \dots\dots(6)$$

The velocity field can be divided into an irrotational component, \mathbf{q}_1 , and a component \mathbf{q}_2 which is associated with the vorticity. Then

$$\mathbf{q} = \mathbf{q}_1 + \mathbf{q}_2, \quad \dots\dots(7)$$

where $\mathbf{q}_1 = -\nabla\phi$. Hence $\nabla \cdot \mathbf{q}_1 = \nabla \cdot \mathbf{q}_2 = 0$; $\nabla^2 \mathbf{q}_1 = 0$; $\nabla \times \mathbf{q}_1 = 0$. Now,

$$\frac{1}{\rho} \nabla p = \frac{1}{\rho} \nabla \rho U \frac{\partial \phi}{\partial x} = U \frac{\partial}{\partial x} \nabla \phi = -2k\nu \frac{\partial \mathbf{q}_1}{\partial x}.$$

Therefore

$$\frac{1}{\rho} \nabla p = \nu \left(\nabla^2 - 2k \frac{\partial}{\partial x} \right) \mathbf{q}_1.$$

In order to satisfy (3) it is thus necessary to have

$$\left(\nabla^2 - 2k \frac{\partial}{\partial x} \right) \mathbf{q}_2 = 0, \quad \dots\dots(8)$$

which suggests putting

$$\mathbf{q}_2 = \frac{1}{2k} (\nabla - 2k\mathbf{i})\chi.$$

Since $\nabla \cdot \mathbf{q}_2 = 0$, this gives

$$\left(\nabla^2 - 2k \frac{\partial}{\partial x} \right) \chi = 0, \quad \dots\dots(9)$$

and the expression for \mathbf{q}_2 is easily shown to satisfy equation (8). Combining this result with (7) gives

$$\mathbf{q} = -\nabla\phi + \left(\frac{1}{2k} \nabla - \mathbf{i} \right) \chi, \quad \dots\dots(10)$$

where ϕ is a solution of Laplace's equation (6) and χ is a solution of (9).

For transverse flow past an infinite cylinder of radius a it is therefore necessary to solve the following equations:

$$\left. \begin{aligned} u &= -\frac{\partial \phi}{\partial x} + \frac{1}{2k} \frac{\partial \chi}{\partial x} - \chi, \\ v &= -\frac{\partial \phi}{\partial y} + \frac{1}{2k} \frac{\partial \chi}{\partial y}, \end{aligned} \right\} \quad \dots\dots(11)$$

where

$$\left(\nabla^2 - 2k \frac{\partial}{\partial x} \right) \chi = 0 \quad \dots\dots(9)$$

and

$$\nabla^2 \phi = 0. \quad \dots\dots(6)$$

The boundary conditions are

$$u = v = 0 \text{ at } r = \infty; \quad u = -U, \quad v = 0 \text{ at } r = a; \quad \dots\dots(12)$$

where $x = r \cos \theta$, $y = r \sin \theta$ and the origin of coordinates is taken at the centre of the cylinder.

A solution of (9) which is found to be accurate to the first order is

$$\chi = C e^{kx} K_0(kr), \quad \dots\dots(13)$$

where $K_0(kr)$ is a modified Bessel function of the second kind. An expansion for (13), valid for small values of kr only, gives

$$\chi \simeq -C(1 + kx)(\gamma + \ln \frac{1}{2} kr). \quad \dots\dots(14)$$

The general solution of Laplace's equation in two dimensional polar co-ordinates is

$$\phi = A_0 \ln r + B_0 \theta + C_0 \theta \ln r + D_0 + \sum_{l=1}^{\infty} r^{\pm l} (A_l \cos l\theta + B_l \sin l\theta); \quad \dots\dots(15)$$

Lamb selects from this the terms giving

$$\phi = A_0 \ln r + A_1 \frac{\partial}{\partial x} \ln r, \quad \dots\dots(16)$$

which makes $\nabla\phi$ tend to zero as r becomes infinite.

Substitution of the values of ϕ and χ from (14) and (16) into (11) gives for the flow field near the cylinder

$$\left. \begin{aligned} u &= - \left(A_0 + \frac{C}{2k} \right) \frac{\partial}{\partial x} \ln r + \left(\frac{Cr^2}{4} - A_1 \right) \frac{\partial^2}{\partial x^2} \ln r \\ &\quad - \frac{C}{2} \left(\frac{1}{2} - \gamma - \ln \frac{1}{2} kr \right) + Ckx(\gamma + \ln \frac{1}{2} kr), \\ v &= - \left(A_0 + \frac{C}{2k} \right) \frac{\partial}{\partial y} \ln r + \left(\frac{Cr^2}{4} - A_1 \right) \frac{\partial^2}{\partial x \partial y} \ln r. \end{aligned} \right\} \quad \dots\dots(17)$$

The first term in each of these expressions is equal to zero. Lamb tacitly neglects the last term in the expression for u , although it is of the first order in kx , and applies the boundary conditions at $r=a$ to obtain

$$A_0 = -\frac{C}{2k}, \quad A_1 = \frac{Ca^2}{4}, \quad C = \frac{2U}{\frac{1}{2} - \gamma - \ln \frac{1}{2} ka}. \quad \dots\dots(18)$$

This solution obeys the surface conditions but begins to fail at a small distance giving velocities which tend linearly to infinity as the distance increases, on account of the approximation (14) introduced for χ . If the exact expression is used for χ we obtain from (13)

$$\begin{aligned} \frac{\partial \chi}{\partial x} &= Ce^{kx} \left\{ kK_0(kr) + \frac{\partial}{\partial r} K_0(kr) \frac{\partial r}{\partial x} \right\} \\ &= Cke^{kx} \left\{ K_0(kr) - K_1(kr) \frac{\partial r}{\partial x} \right\}, \end{aligned}$$

where

$$K_1(kr) = -\frac{\pi}{2} H_1^{(1)}(ikr) = -\frac{\pi}{2} H_1^{(2)}(-ikr).$$

Hence

$$\left. \begin{aligned} \frac{1}{2k} \frac{\partial \chi}{\partial x} - \chi &= -\frac{C}{2} e^{kx} \left\{ K_0(kr) + \frac{x}{r} K_1(kr) \right\}, \\ \frac{1}{2k} \frac{\partial \chi}{\partial y} &= -\frac{C}{2} e^{kx} \frac{y}{r} K_1(kr). \end{aligned} \right\} \quad \dots\dots(19)$$

The calculation of u and v using these equations, together with the value of ϕ given by (16) and the constants (18), leads to a velocity field which is correct at infinity but which does not have the right values on the surface of the cylinder, owing to the neglect of the term $Ckx(\gamma + \ln \frac{1}{2} kr)$ in equation (17). To obtain an accurate result it is necessary to take more terms in the solution of $\nabla^2\chi = 0$.

It is clear that expansion of the exponential in (19) gives terms in x^n in the equation for u and x^ny in the equation for v ; $\partial\phi/\partial x$ and $\partial\phi/\partial y$ must contain corresponding terms, so that the constants can be determined from the conditions at $r=a$. The terms in θ and $\sin \theta$ in the general solution (15) for ϕ are unwanted.

The pressure distribution about the cylinder is obtained by differentiating (15) and using equation (5). It will be seen that terms in (15) containing θ and $\sin l\theta$ lead to odd functions in the pressure equation, so that the pressure distribution would be asymmetrical; this is the type of solution which would be required for a rotating cylinder with which we are not at present concerned. Clearly, also, only negative values of l will be of interest; it was found necessary, after some trials, to go as far as $l = -4$ so as to be sure of the error involved by neglecting higher terms. Equation (16) is therefore replaced by the following expression:

$$\phi = A_0 \ln r + A_1 \frac{\cos \theta}{r} + A_2 \frac{\cos 2\theta}{r^2} + A_3 \frac{\cos 3\theta}{r^3} + A_4 \frac{\cos 4\theta}{r^4}.$$

Differentiation and suitable arrangement of the terms then leads to

$$\left. \begin{aligned} \frac{\partial \phi}{\partial x} &= \left(\frac{A_1}{r^2} - \frac{3A_3}{r^4} \right) + \left(\frac{A_0}{r^2} + \frac{6A_2}{r^4} - \frac{20A_4}{r^6} \right) x + \left(-\frac{2A_1}{r^4} + \frac{24A_3}{r^6} \right) x^2 \\ &\quad + \left(-\frac{8A_2}{r^6} + \frac{80A_4}{r^8} \right) x^3 - \frac{24A_3}{r^8} x^4 - \frac{64A_4}{r^{10}} x^5, \\ \frac{\partial \phi}{\partial y} &= \left(\frac{A_0}{r^2} + \frac{2A_2}{r^4} - \frac{4A_4}{r^6} \right) y + \left(-\frac{2A_1}{r^4} + \frac{12A_3}{r^6} \right) xy \\ &\quad + \left(-\frac{8A_2}{r^6} + \frac{48A_4}{r^8} \right) \left(x^2 y - \frac{24A_3}{r^8} x^3 y - \frac{64A_4}{r^{10}} x^4 y \right). \end{aligned} \right\} \dots\dots(20)$$

Meanwhile, expanding the exponentials in (19) as far as x^3 gives

$$\left. \begin{aligned} \frac{1}{2k} \frac{\partial \chi}{\partial x} - \chi &= -\frac{C}{2} \left\{ K_0 + \left(kK_0 + \frac{K_1}{r} \right) x + \left(\frac{1}{2} k^2 K_0 + \frac{kK_1}{r} \right) x^2 + \left(\frac{1}{8} k^3 K_0 + \frac{k^2 K_1}{2r} \right) x^3 \right\} \\ \frac{1}{2k} \frac{\partial \chi}{\partial y} &= -\frac{C}{2} \left\{ \frac{K_1}{r} y + \frac{kK_1}{r} xy + \frac{k^2 K_1}{2r} x^2 y \right\} \end{aligned} \right\},$$

where K_0 and K_1 are used as abbreviations for the modified second kind Bessel functions of kr . Combining these results and dropping the coefficients of terms of order above x^3 ,

$$\left. \begin{aligned} u &= \left(-\frac{A_1}{r^2} + \frac{3A_3}{r^4} - \frac{CK_0}{2} \right) + \left(-\frac{CkK_0}{2} - \frac{CK_1}{2r} - \frac{A_0}{r^2} - \frac{6A_2}{r^4} + \frac{20A_4}{r^6} \right) x \\ &\quad + \left(-\frac{Ck^2 K_0}{4} - \frac{CkK_1}{2r} + \frac{2A_1}{r^4} - \frac{24A_3}{r^6} \right) x^2 \\ &\quad + \left(-\frac{Ck^3 K_0}{12} - \frac{Ck^2 K_1}{4r} + \frac{8A_2}{r^6} - \frac{80A_4}{r^8} \right) x^3, \\ v &= \left(-\frac{CK_1}{2r} - \frac{A_0}{r^2} - \frac{2A_2}{r^4} + \frac{4A_4}{r^6} \right) y + \left(-\frac{CkK_1}{2r} + \frac{2A_1}{r^4} - \frac{12A_3}{r^6} \right) xy \\ &\quad + \left(-\frac{Ck^2 K_1}{4r} + \frac{8A_2}{r^6} - \frac{48A_4}{r^8} \right) x^2 y. \end{aligned} \right\} \dots\dots(21)$$

Inserting the boundary conditions at $r=a$ and now using K_0 and K_1 as abbreviations for the functions of ka we find

$$\left. \begin{aligned} A_0 &= -\frac{Ca}{2} \left(K_1 - \frac{ka}{4} K_0 + \frac{k^2 a^2}{16} K_1 - \frac{k^3 a^3}{32} K_0 \right), \\ A_1 &= \frac{Ca^2}{4} \left(ka K_1 - \frac{k^2 a^2}{2} K_0 \right), \\ A_2 &= \frac{Ca^3}{32} k^2 a^2 \left(K_1 - \frac{ka}{2} K_0 \right), \\ A_3 &= -\frac{Ca^4}{48} k^2 a^2 K_0, \\ A_4 &= \frac{Ca^5}{32} ka \left(K_0 + \frac{ka}{4} K_1 - \frac{k^2 a^2}{8} K_0 \right) \text{ (from the } x \text{ and } y \text{ terms),} \\ C &= \frac{2U}{K_0 + \frac{ka}{2} \left(K_1 - \frac{ka}{4} K_0 \right)}. \end{aligned} \right\} \dots\dots(22)$$

It is impossible from the coefficients of equation (21) to find a unique value of A_4 . Further expansion of the expression for ϕ would be necessary to fix A_4 , while higher order terms would appear in the other constants. The coefficients of x^3 and $x^2 y$ are out of balance by a quantity of order of magnitude ka .

Since

$$K_0(ka) = -(\gamma + \ln \frac{1}{2} ka) \left(1 + \frac{k^2 a^2}{4} \right) + \frac{k^2 a^2}{4} + O(k^4 a^4),$$

$$K_1(ka) = -\frac{1}{k} \frac{d}{da} K_0(ka) = \frac{1}{ka} + \frac{ka}{2} (\gamma + \ln \frac{1}{2} ka - \frac{1}{2}) + O(k^3 a^3),$$

we can clear the constants of Bessel functions and obtain

$$\left. \begin{aligned} A_0 &= -\frac{Ca}{2} \left\{ \frac{1}{ka} + \frac{3ka}{4} (\gamma + \ln \frac{1}{2} ka - \frac{1}{2}) + O(k^3 a^3) \right\}, \\ A_1 &= \frac{Ca^2}{4} \left\{ 1 + \frac{k^2 a^2}{4} (\gamma + \ln \frac{1}{2} ka - \frac{1}{2}) + O(k^4 a^4) \right\}, \\ A_2 &= \frac{Ca^3}{32} ka + O(k^3 a^3), \\ A_3 &= \frac{Ca^4}{48} k^2 a^2 (\gamma + \ln \frac{1}{2} ka) + O(k^4 a^4), \\ A_4 &= -\frac{Ca^5}{32} ka (\gamma + \ln \frac{1}{2} ka - \frac{1}{2}) + O(k^3 a^3), \\ C &= \frac{-2U}{(\gamma + \ln \frac{1}{2} ka - \frac{1}{2}) - \frac{k^2 a^2}{8} (\gamma + \ln \frac{1}{2} ka + 1)} + O(k^3 a^3). \end{aligned} \right\} \dots\dots(23)$$

The Reynolds' number R is equal to $2Ua/\nu = 4ka$. In view of the fundamental limitation of Oseen's approximation, there is no point in calculating the constants with allowance for terms above $k^3 a^3$. Tables of K_0 and K_1 are given by G. N. Watson (1922) and it is probably easier to use these for computation than it is to use the expansions.

It will be noticed that A_0 , A_2 and A_4 do not reduce to Lamb's values ($-C/2k$, 0 and 0) when only first powers of ka are retained, although A_1 , A_3 and C do so.

Substitution of the values of the constants given by (22) into equation (20), while dropping terms in k^3a^3 and the coefficients of x^4 and x^3y , gives as the solution of the problem, after using (11) and (19):

$$\left. \begin{aligned} \frac{U}{u} \left\{ K_0 + \frac{ka}{2} \left(K_1 - \frac{ka}{4} K_0 \right) \right\} &= -\frac{a^2}{2r^2} \left\{ kaK_1(ka) - \frac{k^2a^2}{2} [K_0(ka)] \left(1 - \frac{a^2}{2r^2} \right) \right\} \\ &+ \frac{ax}{r^2} \left\{ K_1(ka) - \frac{ka}{4} [K_0(ka)] \left(1 - 5 \frac{a^4}{r^4} \right) + \frac{k^2a^2}{16} [K_1(ka)] \left(1 - 6 \frac{a^2}{r^2} + 5 \frac{a^4}{r^4} \right) \right\} \\ &+ \frac{a^2x^2}{r^4} \left\{ kaK_1(ka) - \frac{k^2a^2}{2} [K_0(ka)] \left(1 - 2 \frac{a^2}{r^2} \right) \right\} \\ &+ \frac{a^3x^3}{r^6} \cdot \frac{k^2a^2}{2} K_1(ka) - e^{kx} \left\{ K_0(kr) + \frac{x}{r} K_1(kr) \right\}, \\ \frac{v}{U} \left\{ K_0 + \frac{ka}{2} \left(K_1 - \frac{ka}{4} K_0 \right) \right\} &= \frac{ay}{r^2} \left\{ K_1(ka) - \frac{ka}{4} [K_0(ka)] \left(1 - \frac{a^4}{r^4} \right) \right. \\ &+ \frac{k^2a^2}{16} [K_1(ka)] \left(1 - 2 \frac{a^2}{r^2} + \frac{a^4}{r^4} \right) \left. \right\} \\ &+ \frac{a^2xy}{r^4} \left\{ kaK_1(ka) - \frac{k^2a^2}{2} [K_0(ka)] \left(1 - \frac{a^2}{r^2} \right) \right\} \\ &+ \frac{a^3x^2y}{r^6} \cdot \frac{k^2a^2}{2} K_1(ka) - e^{kx} \frac{y}{r} K_1(kr). \end{aligned} \right\} \dots\dots(24)$$

These formulae fit the boundary conditions closely. Very near the surface of the cylinder there will be some error in the field which is of the same order as that introduced by Oseen's treatment of the dynamical equations and is negligible at small Reynolds' numbers. The expressions can be written in terms of R , neglecting R^2 , as follows:

$$\left. \begin{aligned} \frac{u}{U} \left\{ K_0 \left(\frac{R}{4} \right) + \frac{1}{2} \right\} &= -\frac{a^2}{2r^2} + \frac{ax}{r^2} \left\{ K_1 \left(\frac{R}{4} \right) - \frac{R}{16} \left[K_0 \left(\frac{R}{4} \right) \right] \left(1 - 5 \frac{a^4}{r^4} \right) \right. \\ &+ \frac{R}{64} \left(1 - 6 \frac{a^2}{r^2} + 5 \frac{a^4}{r^4} \right) \left. \right\} + \frac{a^2x^2}{r^4} + \frac{a^3x^3}{r^6} \cdot \frac{R}{8} \\ &- \exp \left(\frac{Rx}{4a} \right) \left\{ K_0 \left(\frac{R}{4} \frac{r}{a} \right) + \frac{x}{r} K_1 \left(\frac{R}{4} \frac{r}{a} \right) \right\}, \\ \frac{v}{U} \left\{ K_0 \left(\frac{R}{4} \right) + \frac{1}{2} \right\} &= \frac{ay}{r^2} \left\{ K_1 \left(\frac{R}{4} \right) - \frac{R}{16} \left[K_0 \left(\frac{R}{4} \right) \right] \left(1 - \frac{a^4}{r^4} \right) \right. \\ &+ \frac{R}{64} \left(1 - 2 \frac{a^2}{r^2} + \frac{a^4}{r^4} \right) \left. \right\} + \frac{a^2xy}{r^4} + \frac{a^3x^2y}{r^6} \cdot \frac{R}{8} \\ &- \exp \left(\frac{Rx}{4a} \right) \frac{y}{r} K_1 \left(\frac{R}{4} \frac{r}{a} \right). \end{aligned} \right\} \dots\dots(25)$$

Velocities computed from equation (25) agree very closely with those from the more complicated expressions when the Reynolds' number is below 0.2. At $R=0.4$ both formulae give a reverse flow near the rear stagnation point due to formation of trailing vortices. The upper limit for the validity of (25) is probably about $R=0.2$. There may be some advantage in using (24) to calculate the field at a distance for higher Reynolds' numbers.

In order to find the drag on the cylinder, W , Filon's (1927) theorem may be used. This shows that

$$W = \rho U E, \quad \dots\dots(26)$$

where E is the outflow of fluid over an imaginary closed surface surrounding the cylinder at a great distance.

Let q_n be the normal outward velocity resolute at a point on a concentric cylindrical surface of radius r , where r is large. Then the net flow of fluid across unit length of this cylinder is

$$\int_0^{2\pi} q_n r d\theta, \quad \dots\dots(27)$$

which must be equal to zero. In evaluating the integral a positive term, E , denoting the outflow will be balanced by a negative term which gives the inflow along the wake.

If we take equation (24) and neglect terms in $1/r$ while putting

$$G = K_0 + \frac{ka}{2} \left(K_1 - \frac{ka}{4} K_0 \right); \quad H = K_1 - \frac{ka}{4} K_0 + \frac{k^2 a^2}{16} K_1,$$

these being functions of ka only, we find

$$\left. \begin{aligned} \frac{ru}{U} G &= aH \cos \theta - re^{kx} \{ K_0(kr) + \cos \theta K_1(kr) \}, \\ \frac{rv}{U} G &= aH \sin \theta - re^{kx} \sin \theta K_1(kr). \end{aligned} \right\}$$

Hence

$$q_n = r(u \cos \theta + v \sin \theta) = \frac{U}{G} \{ aH - re^{kx} [\cos \theta K_0(kr) + K_1(kr)] \}.$$

When r is large

$$\begin{aligned} K_0(kr) &= \left(\frac{\pi}{2kr} \right)^{\frac{1}{2}} e^{-kr}, \\ K_1(kr) &= [K_0(kr)] \left(1 + \frac{1}{2kr} \right) \simeq K_0(kr), \end{aligned}$$

so that

$$q_n r = \frac{2v}{G} \left\{ kaH - \left(\frac{\pi}{2} kr \right)^{\frac{1}{2}} \exp \{ -kr(1 - \cos \theta) \} (1 + \cos \theta) \right\}. \quad \dots\dots(28)$$

When integrated over all values of θ the first term of this equation is the outflow, E , and the second term the inflow. We notice, first, that the part of the second term inside the scroll brackets is independent of the radius a . If powers of ka higher than the first are neglected we have $kaH=1$, but if higher powers are included kaH is a function of a . It is therefore impossible to balance the outflow and inflow unless ka is small. It was not apparent from previous reasoning

that an approximation of this nature had been introduced at a distance from the cylinder. If a more general solution to equation (9) had been taken in place of (13), involving higher order Bessel functions, additional constants would have been included which were functions of a , so that the outflow and inflow terms would presumably have balanced with larger values of ka .

Next it will be observed that along the x axis, behind the cylinder where $\theta = 0$, the inflow term tends to infinity. When θ is sufficiently large, however, between 0 and $\pi/2$, the inflow vanishes as r increases. Thus the inflow region of the wake decreases in width as we go downstream but the inflow velocity rises.

The outflow term gives, using (27),

$$E = 4\pi\nu kaH/G. \quad \dots\dots (29)$$

From (26), therefore,

$$W = 4\pi\eta UkaH/G.$$

Neglecting k^2a^2 etc., we obtain

$$W = \frac{4\pi\eta U}{K_0(ka) + \frac{1}{2}} = \frac{4\pi\eta U}{\frac{1}{2} - \gamma - \ln \frac{1}{2}ka} = \frac{4\pi\eta U}{2.0022 - \ln R} \quad \dots\dots (30)$$

in agreement with Lamb.

In order to evaluate the inflow term it is necessary to consider the integral

$$\int_{-\pi}^{\pi} \{\exp(kr \cos \theta)\}(1 + \cos \theta) d\theta = \int_{-\pi}^{\pi} \exp(kr \cos \theta) d\theta + \int_{-\pi}^{\pi} \{\exp(kr \cos \theta)\} \cos \theta d\theta.$$

Since

$$I_0(x) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \exp(x \cos \theta) d\theta$$

and

$$I_1(x) = \frac{d}{dx} I_0(x) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \{\exp(x \cos \theta)\} \cos \theta d\theta,$$

the integral is equal to $2\pi\{I_0(kr) + I_1(kr)\}$.

For large values of kr the asymptotic expansions of the Bessel functions show that

$$I_0(kr) = I_1(kr) = e^{kr}/(2\pi kr)^{\frac{1}{2}},$$

hence the integral is equal to $2e^{kr}(2\pi/kr)^{\frac{1}{2}}$ and the inflow, from equation (28), is $4\pi\nu/G$, which is equal to the outflow if k^2a^2 and higher terms are negligible.

ACKNOWLEDGMENT

It gives the author very great pleasure to express his thanks to Mr. A. C. Stevenson for his patient and friendly instruction in hydrodynamics and for the trouble he has taken in reading and criticizing this paper.

REFERENCES

- BAIRSTOW, L., CAVE, B. M., and LANG, E. D., 1923, *Phil. Trans. A*, 223, 383.
 FAXÉN, H., 1927, *Nova Acta Soc. Reg. Upsal.* Vol. 1, Ext. Ord.
 FILON, L. N. G., 1927, *Proc. Roy. Soc. A*, 113, 7.
 LAMB, H., 1911, *Phil. Mag.*, 21, 112; 1932, *Hydrodynamics*, 6th Edition (Cambridge: University Press), p. 613.
 WATSON, G. N., 1922, *Theory of Bessel Functions* (Cambridge: University Press).

On the Fluctuating Concentration of X-Ray Products in Water Dispersions

By F. H. KRENZ*

National Research Council, Chalk River, Ontario, Canada

Communicated by N. Feather; MS. received 28th December 1949

ABSTRACT. It has been found that the partial pressure of decomposition gases over water irradiated with x-rays fluctuates more or less regularly with the irradiation time. The experiments suggest that the fluctuations are caused by impurities in the water, and that this phenomenon is possibly related to the periodic rise and fall of electrophoretic mobility observed in the case of certain colloids when these are subjected to x-irradiation.

IN 1937 Crowther and Liebmann announced the discovery of a peculiar effect of x-rays on the electrophoretic mobility of colloidal graphite in water (Crowther and Liebmann 1937). The effect was described as a periodic fluctuation of the electrophoretic mobility of the graphite with the dose of x-rays given to the sol at constant intensity. Liebmann and Jones (see Crowther 1938) showed that the effect could be produced as well in aqueous dispersions of gold. Gray, Read and Liebmann (1941) later showed that the effect could be produced in graphite sols by γ -rays and neutrons.

In the course of investigating the effects of ionizing radiations on water and aqueous systems a phenomenon has been observed in this laboratory which may throw some light on the Crowther effect. It was found that when certain samples of water in a closed, evacuated system were subjected to an intense beam of x-rays the pressure of decomposition gases over the samples rose and fell regularly with time.

The source of radiation in these experiments was a Machlett AEG-50-T x-ray tube operating at full power. The integrated dose rate over the whole sample was determined to be approximately 300 r/sec. The samples consisted of 1 to 3 ml. of water purified by distillation from potassium permanganate and vacuum distilled into a water-jacketed Pyrex cell which admitted x-rays through a thin (0.5 mm.) Pyrex window. The water through the jacket was supplied from a constant temperature bath operating at 15°C.

The pressure was measured with a McLeod gauge after the water vapour was condensed in a cold trap cooled with liquid nitrogen during pressure measurements. The trap and gauge were isolated from the cell during measurements by a stopcock, and the trap was always warmed to room temperature before the stopcock was opened. The measuring technique did not contribute to pressure fluctuations. It was demonstrated in later experiments with a palladium valve that the changes in pressure observed when water was irradiated were due entirely to the evolution of hydrogen or its disappearance.

When the residual pressure of oxygen over the samples was reduced by prolonged pumping to approximately 10^{-4} mm. of mercury (10^{-3} mm. was the limit of the gauge) they could be irradiated for periods up to an hour with a barely detectable increase in pressure (Figure 1, curve A). At residual pressures of 1–5 microns, however, decomposition of the water commenced almost immediately

* Now at the Department of Natural Philosophy, The University, Edinburgh.

upon irradiation. Figure 1, curve B, shows fluctuations in pressure detected with a McLeod gauge during the irradiation of a sample. It was suspected that contamination of the water by mercury might be contributing to this phenomenon, and the McLeod gauge was replaced by a Pirani gauge calibrated for hydrogen. When this was done, and the system baked out *in vacuo* for several hours before the water was distilled into the cell, curve C of Figure 1 was obtained.

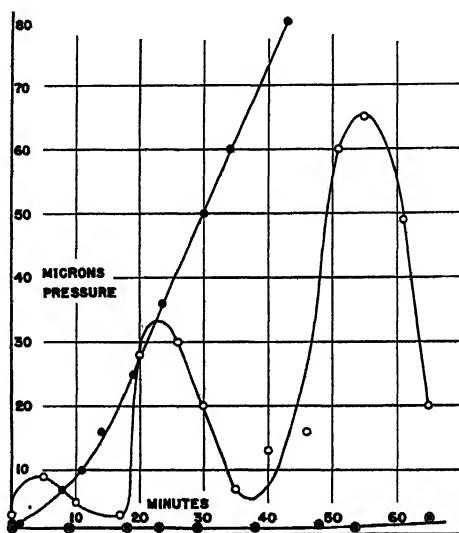


Figure 1. Irradiation of water.

- A: Mercury-contaminated (McLeod gauge), pressure $O_2 \approx 10^{-4}$ mm
- B: Mercury-contaminated (McLeod gauge), pressure $O_2 = 2-3 \mu$.
- C: Mercury-free (Pirani gauge).

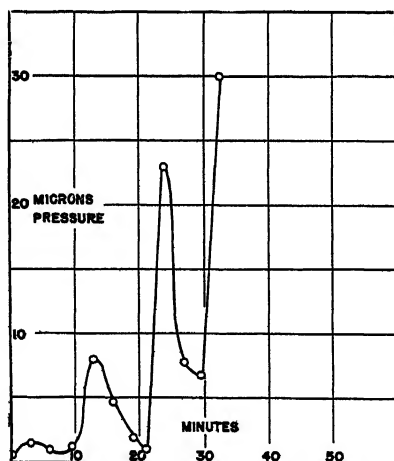


Figure 2. Irradiation of graphite sol.
Temperature of water jacket $15.00 \pm 0.10^\circ \text{C}$.

Figure 2 shows fluctuations in pressure measured with the Pirani gauge over a sample of graphite sol prepared by Crowther's technique (Crowther, Liebmann and Lane 1937). In this case the dose rate was reduced to approximately 100 r/sec. The pressure changes were found to be due entirely to hydrogen.

When the samples of water were purposely contaminated by exposure to mercury vapour and irradiated, pressure fluctuations like those of Figure 1, curve B, could be detected with the Pirani gauge. It was also found that the pressure continued to fluctuate when the x-rays were turned off (Figure 3).

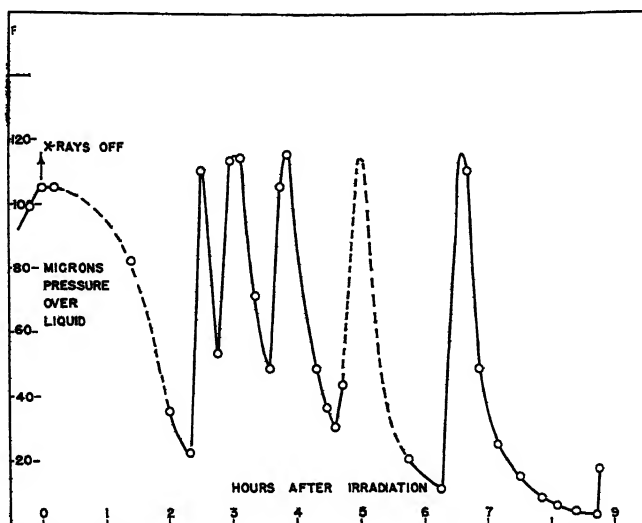


Figure 3. Post-irradiation phenomenon : water (mercury-contaminated).
Temperature of water jacket $15.15 \pm 0.10^\circ \text{C}$.

These fluctuations were maintained at least up to 24 hours after irradiation. The pressure changes were found to be entirely due to hydrogen. No such after-effect was observed in the case of colloidal graphite.

REFERENCES

- CROWTHER, J. A., 1938, *Nature, Lond.*, **142**, 569.
 CROWTHER, J. A., and LIEBMANN, H., 1937, *Nature, Lond.*, **140**, 28.
 CROWTHER, J. A., LIEBMANN, H., and LANE, T. B., 1937, *Phil. Mag.*, **24**, 654.
 GRAY, L. H., READ, J., and LIEBMANN, H., 1941, *Brit. J. Radiol.*, **14**, 102.

On the Retarding Field Current in Diodes

BY R. FÜRTH* AND D. K. C. MACDONALD†

MS. received 26th October 1949

ABSTRACT. In a previous paper the authors have shown how the space charge distribution in plane and cylindrical diodes under retarding field conditions can be calculated by well-known methods of statistical thermodynamics, and have derived expressions for the upper limit of current for which this condition holds. In the present paper certain inaccuracies in the previous treatment are rectified, the argument is completed, and certain erroneous views, expressed in a recent paper by D. A. Bell, are criticized.

IN a paper on the spontaneous fluctuations of electricity in thermionic valves under retarding field conditions (MacDonald and Fürth 1947) we have shown that the distribution of space-charge density and the upper limit I_c of the current for which true retarding field conditions hold can be obtained in a comparatively simple way by the application of well-known methods of statistical thermodynamics. In the present paper we wish to rectify certain inaccuracies in the previous treatment and to complete the argument, and finally once again to stress the significance of the results for the determination of cathode temperatures in thermionic valves in view of certain erroneous views expressed in a recent paper by D. A. Bell (1949).

In the case of a plane-parallel diode the maximum ρ_m of the space-charge density, which under the above-mentioned limiting condition must coincide with the cold electrode, is equal to

$$\rho_m = \frac{\pi k T}{8 e d^2}, \quad \dots \dots (1)$$

where e is the magnitude of the electronic charge and d the inter-electrode distance. The current I is connected with the saturation current J by

$$I = J \exp \left[\frac{e(V + v_a)}{kT} \right], \quad \dots \dots (2)$$

where V is the (negative) anode potential and v_a a possible potential barrier at the surface of the cold electrode. The saturation current satisfies the classical Richardson relation

$$J = \rho_0 a (kT/2\pi m)^{1/2}, \quad \dots \dots (3)$$

in which ρ_0 is the electronic density at the hot-electrode surface and m the electronic mass. Finally, one has for thermodynamic equilibrium within the space charge

$$\rho_m = \rho_0 \exp [eV/kT]. \quad \dots \dots (4)$$

From (1), (2), (3) and (4) one obtains the expression for the limiting current I_c

$$I_c = \frac{\sqrt{\pi}}{8\sqrt{2}} \frac{k^{3/2}}{e\sqrt{m}} \frac{T^{3/2}a}{d^2} \exp [ev_a/kT], \quad \dots \dots (5)$$

which is identical with equation (31) of our first paper, and which differs only by an insignificant numerical factor from the expression obtained on the basis of Langmuir's kinetic theory.

* Birkbeck College, University of London.

† Clarendon Laboratory, University of Oxford.

In the case of a cylindrical diode equation (1) is to be replaced by

$$\rho_m = \frac{1+B^2}{(2\pi e/kT)r_a^2}, \quad \dots\dots (6)$$

where r_a is the radius of the anode-cylinder and B is the root of the transcendental equation

$$\tan[B \log(r_a/r_c)] = -B, \quad \dots\dots (7)$$

r_c being the radius of the cathode cylinder.

Instead of the relation (2) we have now to use an expression obtained by Schottky (1914) from elementary dynamical considerations:

$$I = \frac{2}{\sqrt{\pi}} J \left\{ e^{-A} \int_0^{\sqrt{(A/\eta)}} \exp[(\eta-1)x^2] dx + \int_{\sqrt{(A/\eta)}}^{\infty} \exp(-x^2) dx \right\}, \quad \dots\dots (8)$$

$$\text{with} \quad \eta = 1 - (r_c/r_a)^2 \quad \dots\dots (9)$$

$$\text{and} \quad A = -eV/kT. \quad \dots\dots (10)$$

The expression (8) has to be multiplied by the factor $\exp(e v_n/kT)$ in the case of the existence of a potential barrier v_n .

For $\eta \ll 1$, (8) becomes

$$I = \frac{r_a}{r_c} J e^{-A}, \quad \dots\dots (8a)$$

and for $\eta \simeq 1$ and $A \gg 1$ it can be expanded in the form

$$I = \frac{2}{\sqrt{\pi}} J e^{-A} \sqrt{A} \left[1 + \frac{1}{2A} - \frac{1}{4A^2} + \dots \right]. \quad \dots\dots (8b)$$

In any case I always contains the Boltzmann factor $\exp(-A) = \exp(eV/kT)$, in agreement with general results of statistical thermodynamics, whereas a factor of the form $\exp(-A/2)$, as suggested by Bell (1949), is clearly in contradiction to statistical thermodynamics.

The limiting current is obtained from (6) and (8) in connection with (3) and (4), which remain in force. Thus we obtain for a cylinder of length l

$$I_c = \frac{\sqrt{2}(1+B^2)}{\pi} \frac{k^{3/2}}{e\sqrt{m}} \frac{T^{3/2} l r_c}{r_a^2} \left\{ \int_0^{\sqrt{(A/\eta)}} \exp[(\eta-1)x^2] dx + e^A \int_{\sqrt{(A/\eta)}}^{\infty} \exp(-x^2) dx \right\} \\ \times \exp\left(\frac{e v_n}{kT}\right), \quad \dots\dots (11)$$

which has to replace equation (36) of our first paper.

The statistical equilibrium of the space charge, which is the basic assumption of our treatment, is, strictly speaking, incompatible with the flow of a steady current. Consequently the formula (4) and the expressions (1) and (6) are not strictly valid (especially in the neighbourhood of the minimum), but the relations which are obtained by eliminating ρ_m from (1) and (4) or (6) respectively are still very nearly correct. This accounts for the fact that the resulting expressions for I_c coincide essentially with those obtainable from detailed kinetic treatment. It also indicates why it would in fact be wrong to attempt to evaluate I_c directly as the 'emission' current from the surface $\rho = \rho_m$ from formula (3) by replacing ρ_0 by ρ_m and identifying a with the area of that surface. This procedure happens to give the right result in the plane-parallel case because of the non-diverging character of the current flow, but would give a wrong result in the cylindrical case, in particular yielding a finite value for I_c for $r_c \rightarrow 0$.

When r_a/r_c is of the order of magnitude unity it follows from (7) that

$$B \sim \frac{\pi}{2 \log(r_a/r_c)} \gg 1 \quad \dots\dots (12)$$

and, further, $\eta \ll 1$. Thus (11) becomes

$$I_c = \frac{\pi^{3/2}}{4\sqrt{2}} \frac{k^{3/2}}{e\sqrt{m}} \frac{T^{3/2}l}{r_a \log^2(r_a/r_c)} \exp\left[\frac{ev_a}{kT}\right] \left(\frac{r_a}{r_c} \sim 1\right). \quad \dots\dots (13)$$

This differs from our previous expression (37) by a factor r_a/r_c , and is now, apart from the exponential factor involving v_a , identical with the corresponding formula of Möller and Detels (1926) obtained from kinetic theory under the condition $B \gg 1$.

When on the other hand r_a/r_c is large compared with unity, one has from (7)

$$B \sim \frac{\pi}{\log(r_a/r_c)}, \quad \dots\dots (14)$$

and this is evidently not large compared with unity. Hence the application of formula (13) to this case by Möller and Detels is unjustified. From (4), (6) and (10) follows

$$A = \log(\rho_0/\rho_m) = \log\left[\frac{2\pi e\rho_0 r_a^2}{kT(1+B^2)}\right], \quad \dots\dots (15)$$

which under ordinary conditions is a large quantity. Hence one can use (8*b*), so that (11) becomes

$$I_c = \frac{\sqrt{2}}{\pi} (1+B^2) \frac{k^{3/2}}{e\sqrt{m}} \frac{T^{3/2}l}{r_a^2} \sqrt{A} \left[1 + \frac{1}{2A} - \frac{1}{4A^2} + \dots\right] \exp\left(\frac{ev_a}{kT}\right) \left(\frac{r_a}{r_c} \gg 1\right). \quad \dots\dots (16)$$

For a finite fixed value of r_a , the radius of the anode cylinder, under this condition r_c is very small, i.e. the cathode becomes a thin filament, and the expression (16) tends towards zero unless ρ_0 is concomitantly made very large, so that the quantity $r_c\sqrt{(\log \rho_0)}$ remains finite. For a finite fixed value of r_c , on the other hand, r_a becomes very large, and as A only increases logarithmically with r_a whilst the denominator is proportional to r_a^2 , we again obtain a vanishingly small limiting current.

It thus follows that, unless one wishes to resort to extremely small currents, one must maintain $r_a/r_c \sim 1$ in experimental investigations of the retarding field region, in particular directed towards the determination of cathode temperatures. If, however, experiments are conducted *above* the limiting current, a spurious (and *too large*) apparent cathode temperature will ensue from slope measurements of the characteristic. This holds specifically for the experiments quoted by Bell. Also, as pointed out in our previous paper, the method of logarithmic plotting can easily mask a slow 'power' variation of one of the variables involved.

In view of all these considerations the method of plotting the product IR_a (where R_a is the differential resistance of the valve) against I for currents below I_c , as explained in §5 of our first paper, which does not suffer from the above-mentioned sources of error, appears to be the most adequate method for such investigation.

REFERENCES

- BELL, D. A., 1949, *Proc. Phys. Soc. B*, **62**, 334.
 MACDONALD, D. K. C., and FÜRTH, R., 1947, *Proc. Phys. Soc.*, **59**, 375.
 MÖLLER, H. G., and DETELS, F., 1926, *Žb. drahtl. Telegr. Teleph.*, **27**, 74.
 SCHOTTKY, W., 1914, *Ann. Phys., Lpz.*, **44**, 1011.

CONTENTS FOR SECTION A

	PAGE
Dr. E. MARSDEN. Rutherford Memorial Lecture (1948)	305
Mr. R. W. CAHN. A New Theory of Recrystallization Nuclei	323
Dr. E. E. SALPETER. Nuclear Induction Signals for Long Relaxation Times	337
Dr. C. B. VAN WYK. γ -Decay and Mesons of Spin Zero	350
Mr. C. BALAKRISHNAN and Dr. J. D. CRAGGS. On the Velocity of Discharge Propagation in Externally Quenched Geiger Counters	358
Prof. C. A. COULSON and Mr. N. H. MARCH. Momenta in Atoms using the Thomas-Fermi Method	367
Prof. L. F. BATES and Mr. F. E. NEALE. A Quantitative Study of the Domain Structure of Single Crystals of Silicon-Iron by the Powder Pattern Technique.	374
Dr. J. C. WILLMOTT. The Infra-red Spectrum of Magnesium Oxide	389
Letters to the Editor:	
Mr. F. K. GOWARD, Mr. V. L. TELEGDI and Mr. J. J. WILKINS. The Photo-disintegration of Carbon into Three Alpha-particles	402
Mr. D. BIJL. Paramagnetic Resonance Absorption in Titanium Caesium Alum at Low Temperatures	405
Dr. B. BLEANEY. Anisotropy in Titanium Alum	407
Dr. B. BLEANEY and Mr. D. J. E. INGRAM. Paramagnetic Resonance in Copper Fluosilicate	408
Mr. A. ABRAGAM and Prof. M. H. L. PRYCE. Theoretical Interpretation of Copper Fluosilicate Spectrum	409
Dr. J. K. MACKENZIE and Prof. N. F. MOTT. A Note on the Theory of Melting	411
Mr. R. K. LAIRD and Dr. R. F. BARROW. An Ultra-violet Band-system of CS_2^+	412
Reviews of Books	413
Contents for Section B	415
Abstracts for Section B	416

ABSTRACTS FOR SECTION A

Rutherford Memorial Lecture (1948), by E. MARSDEN.

A New Theory of Recrystallization Nuclei, by R. W. CAHN.

ABSTRACT. After a review of the numerous experimental observations in this field which need to be explained, a new theory is presented of nucleus formation when a cold-worked metal is annealed. The nuclei are held to be formed in the most distorted parts of the lattice, in the 'local curvatures' postulated by Burgers. At these places the lattice locally regains its perfection by the recently discovered process of diffusion of dislocations, or 'polygonization', and the perfect nucleus so created is then able to consume the less perfect crystal in its vicinity. Experimental evidence for this picture is presented. Each potential nucleus is held to have a definite incubation period (as distinct from the fixed probability of becoming active which was the basis of previous theories). By means of certain assumptions as to the distribution of curvatures and the dependence of incubation time on curvature, a quantitative theory is developed, which is found to be capable of accounting for all the experimental data on nucleation rates. The paper concludes by comparing critically the new theory with older ones.

Nuclear Induction Signals for Long Relaxation Times, by E. E. SALPETER.

ABSTRACT. Equations of motion, previously derived by Bloch (1946), for the nuclear magnetic moment in a nuclear induction experiment are applied to the steady state reached when the strong magnetic field H_z is swept sinusoidally about its resonance value. Different approximations for the shapes and sizes of nuclear induction signals are derived for various limiting cases, depending (a) whether the passage of H_z is 'adiabatic' or 'extremely non-adiabatic' (very weak radio-frequency field) and (b) whether the relaxation times are shorter or longer than the period of the sweep of H_z . It is found that quite large signals may be induced by even a very weak radio-frequency field if the relaxation times are very long.

 γ -Decay and Mesons of Spin Zero, by C. B. VAN WYK.

ABSTRACT. On the assumption that τ -mesons are scalar particles and have an appreciable interaction with nucleons, the lifetime of a τ -meson at rest decaying into two photons and a pseudoscalar π -meson, is calculated by means of standard perturbation theory. The lifetime obtained is $G \times 10^{-12}$ sec. where G is the reciprocal of the product of the dimensionless meson-nucleon coupling parameters. This result is roughly what can be expected from observations and, taken together with the results of a previous paper, seems to indicate that the heavy mesons have spin zero.

On the Velocity of Discharge Propagation in Externally Quenched Geiger Counters, by C. BALAKRISHNAN and J. D. CRAGGS.

ABSTRACT. A new method for measuring the propagation velocity of the discharge along a Geiger non-proportional counter is described. The results obtained with cylinder grade argon, hydrogen and neon are summarized graphically, and a brief discussion of propagation mechanisms is given.

Momenta in Atoms using the Thomas-Fermi Method, by C. A. COULSON AND N. H. MARCH.

ABSTRACT. The Thomas-Fermi statistical theory of the free atom is used to calculate the momentum distribution and the shape of the Compton profile for x-ray scattering, the results being expressed in dimensionless form applicable to all atoms. The mean momentum, which varies as $Z^{2/3}$, where Z is the atomic number, agrees well with more accurate wave-mechanical calculations. Numerical results are also reported for argon and krypton using the Dirac modification of the original theory, in which exchange effects are taken into account and the atom has a finite radius. Agreement with experiment is not noticeably better with these refinements.

A Quantitative Study of the Domain Structure of Single Crystals of Silicon-Iron by the Powder Pattern Technique, by L. F. BATES AND F. E. NEALE.

ABSTRACT. A quantitative study of the domain structure of single crystals of silicon-iron has been made by the powder pattern method, special examination being made of the patterns formed on a strip specimen whose surface was approximately a (100) plane with a [011] direction parallel to the long edge of the strip. For Mode III magnetization, which occurs when a field acts inside the specimen parallel to the [011] direction, the line deposits formed at right angles thereto were on the average spaced in good accord with Néel's theory in the case of wide strips, but in less good accord for narrow strips. Qualitative results of other measurements on strip and disc specimens are reported and interpreted.

The Infra-red Spectrum of Magnesium Oxide, by J. C. WILLMOTT.

ABSTRACT. An investigation of the infra-red spectrum of magnesium oxide has been carried out in the range up to 25.5μ , both in reflection and transmission. The eigen-frequency is found to be at $17.5 \pm 0.3 \mu$ and four subsidiary maxima have been found in transmission. The results of Barnes, Brattain and Seitz in the range up to 15.6μ have not been confirmed and no detailed fine structure has been found. The reflection spectrum of magnesium oxide has been fully investigated up to 25.5μ , and a broad reflection maximum has been found extending from 17μ to 21μ , with subsidiary maxima at 14.02μ and 23.8μ .

From the data obtained curves have been drawn to show the behaviour of the two optical constants n and k in the range from 7μ to 25μ .

THE PROCEEDINGS OF THE PHYSICAL SOCIETY

Section B

VOL. 63, PART 5

1 May 1950

No. 365 B

Summarized Proceedings of Symposium on Applications of Ultrasonics

By G. BRADFELD

National Physical Laboratory, Teddington, Middx.

Communicated by the Acoustics Group of the Society; MS. received 26th July 1949

ABSTRACT. A symposium of the Acoustics Group of the Physical Society, held on 18th February 1949, surveyed recent advances in (a) the investigation of the fundamental structure of matter; (b) telecommunication and allied applications; (c) use of mechanical forces set up by intense waves.

In (a) derivation of elastic constants of matter was an important field especially as small samples such as single crystals could be used. Losses incurred in propagating waves were surveyed with the help of an electrical transmission line model and simple versions of this were established to represent relaxation phenomena based on Maxwell's hypothesis of shear elasticity as a time function and on Kneser's treatment of loss due to delay in a storage process. There was excellent agreement of the latter with recent results on acetic acid.

Available sources of ultrasonic power were surveyed and the importance of barium titanate as a powerful and strongly coupled piezoelectric transducer was emphasized.

An expression for the receiver/transmitter power ratio in telecommunications systems was examined for gaseous, liquid and solid media and, from available data, optimum frequencies for various ranges were deduced. These were found to be in accord with experience in echo-sounding, earth exploration and in propagation in metals.

Accounts were given of experience with flaw detectors and echo-sounding which showed that these were becoming important industrially and in navigation; work on blind aids was unpromising. Advances in timing and time delay devices were described.

There is a dearth of important industrial applications of the use of intense waves in spite of the interesting phenomena which have been demonstrated in the laboratory. The importance of the study of cavitation was pointed out. Results were discussed for killing bacteria, disintegrating proteins, emulsifying, soldering aluminium and refining the crystalline structure in solidification of light alloys. Stress was laid on the wideness of the frequency spectrum over which these phenomena occurred and on the difficulties and importance of maintaining temperature constant and of measuring intensities during investigations.

§ 1. THE INVESTIGATION OF THE FUNDAMENTAL STRUCTURE OF MATTER

IN this first section the simplest application is the derivation of elastic constants of solids and of adiabatic compressibility of fluids, through measurement of velocity. In solids an important case is that of small single crystals, the anisotropic properties of which can be more readily measured by ultrasonic waves than by alternative methods. Velocity and absorption tests in gases yield valuable information about molecular shape and spin and collision phenomena. Gaps in

our knowledge exist of the performance of gases at high temperatures and the properties of mixtures of gases. In liquids the recent recognition of the error arising from neglect of dilatational viscosity has removed one serious divergence between theory and experiment.

Dr. Pinkerton emphasized the value of ultrasonic velocity and absorption measurements in determining relaxation times and thus providing information about molecular equilibria not otherwise obtainable. Three classes of relaxation phenomena may be distinguished: (i) the sharing of vibratory energy; (ii) changes in the local order of the molecules affecting the volume they occupy; (iii) chemical or quasi-chemical changes affecting the relative proportions of two or more molecular species.

The first and last effects were thought to be due to the temperature changes involving a relaxing specific heat while the second may be due either to this cause or to the vibratory changes or to both together.

It is convenient to consider the behaviour of a material propagating mechanical waves by using a model in the form of an electrical transmission line propagating an electromagnetic wave, the analogous expressions for the velocities v in these cases being

$$v_e = (LC)^{-\frac{1}{2}}, \quad \dots\dots(1)$$

$$v_s = (\rho/n)^{-\frac{1}{2}} \text{ (shear wave in solid),} \quad \dots\dots(2)$$

$$v_l = \left\{ \frac{\rho}{4n/3 + k} \right\}^{-\frac{1}{2}} \text{ (longitudinal wave in solid),} \quad \dots\dots(3)$$

$$v_l = (\rho\beta)^{-\frac{1}{2}} \text{ (longitudinal wave in liquid),} \quad \dots\dots(4)$$

where L , the inductance per unit length, corresponds with density ρ , and C the capacity per unit length corresponds with $1/n$, the reciprocal of the modulus of rigidity, or with $1/(4n/3 + k)$, or with β , the adiabatic compressibility ($=1/k$) respectively in (2), (3) and (4).

The attenuation of a line with a resistance R in series with C (Figure 1(a)), $R \ll 1/\omega C$, is

$$\alpha = \frac{1}{2}\omega^2 R(LC^3)^{\frac{1}{2}}, \quad \text{i.e. } \alpha/\nu^2 = 2\pi^2 R(LC^3)^{\frac{1}{2}},$$

so that α/ν^2 is independent of frequency ν , and the model correctly represents the attenuation-frequency characteristic common to all fluids in the absence of relaxation phenomena.

Maxwell's hypothesis that viscosity be regarded as a shear elasticity diminishing exponentially with time leads, as suggested by Mason (1947), to Figure 1(b), a condenser C_s shunting R with the normal relationship of α and ν at low frequencies but with a velocity increase to $[(1/C + 1/C_s)/L]^{\frac{1}{2}}$ at high frequencies, when the attenuation is independent of frequency.

Liquids whose molecules are highly polymerized have been studied by Mason and his co-workers (1948). They found that polymers of isobutylene behave like solids at high frequencies in transmitting shearing waves and possessing shear elasticity. The velocity of longitudinal waves increased by 20% or more with increasing frequency. The shear elasticity possessed two components associated with different relaxation times, and the shorter time was found to vary with temperature in a manner similar to that in the case of acetic acid discussed below. These two components of shear elasticity are ascribed to two different types of

relative motion between the long chain-like molecules of the polymers. The theoretical treatment given by Mason is in terms of an equivalent transmission line more complex than any of those discussed above; good agreement between theory and experiment was found.

During compression, energy is shared by various storage processes, so that a definite fraction is involved in any one relaxation process. The transmission

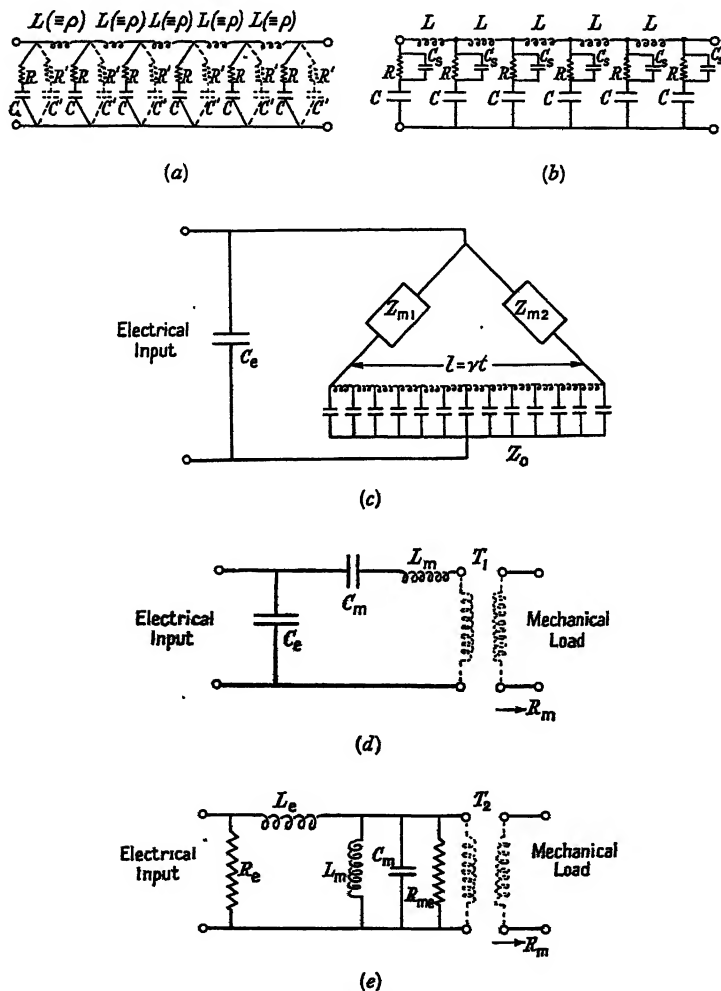


Figure 1.

line model can cover this case by adding to the simple network across R and C an additional pair of series elements R' and C' (see Figure 1(a)), of time constant τ , where $C' \ll C$. It can readily be shown that α , the attenuation constant, is given by

$$\alpha = (C^1/2CV_\infty)\omega^2\tau/(1 + \omega^2\tau^2), \quad \dots\dots(5)$$

where V_∞ is the velocity at high frequencies. Maximum attenuation per wavelength μ_m , where $\mu = \alpha\lambda = \alpha v_1/\omega = 2\pi\alpha v_1/\omega$, occurs when $\tau = 1/\omega$, i.e.

$$\mu_m = \frac{1}{2}\pi C^1/C. \quad \dots\dots(6)$$

Introduction of μ_m in (5) leads to the relation for α/ν^2 proposed by Kneser (1938). Similarly it is clear that at high frequencies such that $R^1 \gg 1/\omega C^1$ the velocity is independent of C^1 so that the total dispersion $(V_\infty - V_0)/V_\infty$ is, as given by Kneser,

$$(V_\infty - V_0)/V_\infty = C^1/2C = \mu_m/\pi. \quad \dots\dots(7)$$

Equation (7) shows that the total molecular energy which is involved in the relaxation process can be found by measuring either μ_m or $(V_\infty - V_0)/V_\infty$; the measurements have necessarily to be carried out in the vicinity of the critical frequency $\omega = 1/\tau$.

In the case of those relaxation processes where the equilibrium is shifted by the adiabatic changes of temperature during propagation, the ratio C^1/C takes the form $(\gamma - 1)Ci/C_p$, where γ is the ratio of the specific heats and Ci is a relaxing specific heat.

Bauer's (1949) treatment of absorption in simple organic liquids with symmetrical molecules and no hydroxyl groupings shows that the abnormally

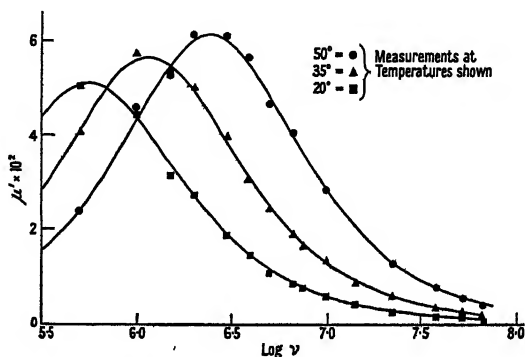


Figure 2. Absorption in acetic acid.

Plot of μ' , the excess of absorption λ over a constant term $A\nu^2$ for frequencies 0.5 to 67 Mc/s. (Lamb and Pinkerton 1949).

Curves were plotted from equation (5), using experimental values of constants.

high value is due to slow exchange of the vibratory energy, consonant with the weak intermolecular forces in these liquids. He predicts such high values for ν_m for typical liquids in this class, e.g. benzene and carbon tetrachloride, that the demonstration of relaxation effects in them would be very difficult in practice. Rapuano (1948), however, finds some evidence of abnormality in α/ν^2 between 10 and 75 Mc/s. in carbon disulphide.

In a class of liquids including water and the lower primary alcohols, the absorption is two or three times that expected from shear viscosity and is related to it, but no evidence for a relaxation process exists. In contrast with the previous class, absorption falls rapidly with temperature (Pinkerton 1949).

A relaxation process apparently arising from chemical dissociation or other change of molecular structure is evident in acetic acid, in aqueous solutions of its salts and in its methyl and ethyl esters. Investigation of acetic acid showed τ to be about 0.32 microsecond at 17°C., and Spakowski (1938) suggested that such a long time of equilibration corresponds to a chemical change connected with the linking and break-up of the dimer. The closeness of theory and experiment is shown in Figure 2, in which the excess absorption per wavelength is plotted against

$\log \nu$; the difference between the experimental points and the curves calculated from equation (5) by Lamb and Pinkerton (1949) is remarkably small.

The rapid change of τ with temperature evident in Figure 1 follows the law

$$\tau = \tau_0 \exp(\Delta E_a/RT),$$

E_a , which is about 8,900 cal/gm. mol, can be regarded as an activation energy (cf. Hall 1948). Should the dissociation of the dimer of acetic acid be responsible for relaxation effects, then from the known heat of dissociation it should be possible to calculate the absorption and dispersion. Since this yields a value of absorption much less than that observed and also predicts a more violent variation of dispersion with temperature, the precise nature of the process remains somewhat obscure.

Dr. Pinkerton felt that there was little doubt that ultrasonic studies of the types outlined above were likely to reveal much that is fundamental about the liquid state.

§ 2. TELECOMMUNICATION AND ALLIED APPLICATIONS

As a preliminary to a review of applications in the fields of telecommunication and of high intensity irradiation, Mr. Bradfield discussed electromechanical transducer performance. The general network for the piezoelectric type is given in Figure 1(c), where Z_{m_1} and Z_{m_2} are proportional to the terminating mechanical impedances on the element represented by a transmission line of travel time l/v , l being the length and v the mechanical wave velocity; near the half-wavelength resonance, this network simplifies to Figure 1(d), where the transformer ratio of T_1 involves the dimensions and piezoelectric constant. The corresponding network for a magnetostrictive transducer is given in Figure 1(e), but in this case the transformer T_2 , besides involving a magnetostrictive constant, also inverts the mechanical load impedance. In Figures 1(d) and (e), the ratios C_m/C_e and L_m/L_e represent the tightness of coupling of the electric and mechanical sides of the transducer and are proportional to the ratio of stored mechanical energy to stored electric (or magnetic) energy. In Table 1 performances are compared for typical materials, including also values for the moving conductor type of transducer.

The values for the hammer blow (i.e. last line on Table 1) obtainable from these materials show that barium titanate should be over 100 times better than quartz or nickel. The coupling factor C_m/C_e , which is a measure of bandwidth, was similarly about 100 times as good as for quartz. Although little detailed work has yet been done on the use of barium titanate for ultrasonic applications, its potentialities for this work seem to be very great.

The maximum intensity from various materials for c.w. operation will approximate to the expression $\Delta^2 E^{\frac{1}{2}}/\rho^{\frac{1}{2}}$, with a value of rather over 60 watts/cm² for a single-ended X-cut quartz transducer, which agrees well with practical experience.

High efficiency, though not of paramount importance in telecommunication applications, may prove vital for industrial applications requiring intense irradiation. Quartz transducers are good and barium titanate* should be similar but much more convenient in use. The efficiency of magnetostrictive transducers is

* Note added in proof. Whether its low Q (150-200) and consequent heating will be troublesome has yet to be seen.

Table 1. Comparison of Ultrasonic Sources

	Note	Piezoelectric		Magnetostriction		Moving conductor	
		Quartz X-cut	Barium Titanate (Mason 1948)	Nickel	Cobalt-Iron (50 : 50)	Copper	Aluminium
Density	ρ	2.66	5.5	8.9	8.2	8.9	2.7
Young's modulus in dynes/cm ²	E	$7.8 \times 10^{11} (=1/S_{11})$	12×10^{11}	20×10^{11}	22×10^{11}	12.3×10^{11}	7.1×10^{11}
Velocity (m/sec.)	$c = (E/\rho)^{1/2}$	5400	4700	4750	5200	3710	5120
Characteristic impedance (c.g.s. units)	$(E\rho)^{1/2}$	1.4×10^8	2.6×10^8	4.23×10^8	4.10×10^8	3.31×10^8	1.38×10^8
Electric or magnetic strain conditions	—	(3)					
Mechanical strain (max.)	Δ	± 130 kv/cm.	+30 kv/cm. to -7.5 kv/cm.	5500 lines/cm ²	20000 lines/cm ²	10000 amp/cm ²	6600 amp/cm ²
Open-circuit pressure	$P_{\max} = E\Delta$	$\pm 3 \times 10^{-8}$	6×10^{-4} total	102 oersteds	66 oersteds	15000 lines/cm ²	15000 lines/cm ²
(dynes/cm ²)	—	47×10^8	720×10^8	3×10^{-5}	4.9×10^{-5}	—	—
(kg/cm ²)	—	48	730	61	107	—	—
Short-circuit velocity	$c\Delta$	(1)					
(cm/sec.)	—	32	280	14	25	$6\frac{1}{2}$	10
Crushing strength (dynes/cm ²)	—	10^{10}	—	5.3×10^9	—	3.5×10^9	1.8×10^9
Maximum intensity (watts/cm ²)	I_{\max}	(1) (2)	5000	21	63	$3\frac{1}{2}$	$3\frac{1}{2}$

(1) Complete reversal for quartz, 30 to -7.5 kv/cm. for barium titanate. Zero to maximum current for moving conductor.

(2) $I_{\max} = \frac{1}{2} \Delta^2 (E^2/\rho)^{1/2}$ into matched load at each end.

(3) 102 oersteds is for half-hard nickel. Decreases to about 20 oersteds for soft nickel.

lower but, for lower frequencies, about 65% efficiency can be attained by good design using nickel radiating into water; a somewhat higher efficiency is possible for cobalt-iron for alloys.

For generating or receiving waves in air, transducers with solid elements are disadvantageous because of the low characteristic impedance of the medium, and although bilamellar elements help, especially for reception, a high power transmitter of the siren type (Allen and Rudneck 1947) is much more promising. The source intensity is here around 400 watt/cm² and efficiency about 20%, while powers up to 2 kw., or even more, do not seem difficult to obtain.

In telecommunication and allied applications, absorption plays as important a part as geometrical spreading in limiting the range of ultrasonic propagation. The overall attenuation of the system is the product of factors representing radiating and receiving array gains, geometrical spreading and absorption loss together with a scattering or reflection coefficient in the case of echoes of radar type. For fluids in the absence of relaxation phenomena the attenuation can be expressed

$$\frac{P_r}{P_g} = \text{const.} \frac{D^4}{\lambda^2 x^2} \exp\left(-\frac{2ax}{\lambda^2}\right), \quad \dots\dots(8)$$

where P_r is the electrical power from the receiving transducer, P_g the electrical power input to the transmitting transducer, $D(\gg \lambda)$ is the diameter of the transmitting and receiving arrays, x is the total path length and $a = \alpha \lambda^2$ and is independent of frequency.

Equation (8) can apply to solids if the index 2 of λ in the exponential term be replaced by m , the value of which varies from unity for material of fine grain at lower frequencies to about 4 for coarser grains and higher frequencies. The index of x shown as 2 applies to extensive specularly reflecting surfaces. For extensive scattering surfaces this index rises to 4.

These expressions lead to optimum wavelengths λ_0 for propagating over any given path length with minimum attenuation:

$$\text{for fluids } (m=2), \quad \lambda_0 = (2ax)^{\frac{1}{2}}, \quad \dots\dots(9)$$

$$\text{for solids } (m=1), \quad \lambda_0 = a'x, \quad \dots\dots(10)$$

where the exponent in (8) is $-2a'x/\lambda$.

Some modification has to be made in the case of echoes from small targets. The optimum wavelength λ_0' then equals about $0.43 \lambda_0$.

In this way various media and various distances have appropriate optimum frequencies as in Table 2.

Table 2

Medium	a or a'	Optimum frequency for given distance				
		1 ft.	10 ft.	100 ft.	1000 ft.	10000 ft.
Nitrogen	1.9×10^{-4}	—	98 kc/s.	31 kc/s.	9.8 kc/s.	3.1 kc/s.
Pure water	5.5×10^{-6}	8 Mc/s.	2.5 Mc/s.	800 kc/s.	250 kc/s.	80 kc/s.
Sea-water	3×10^{-4}	Relaxation region			35 kc/s.	11 kc/s.
Mudstone	$a' = \alpha \lambda = 0.08$ ($m=1$)	120 kc/s.	12 kc/s.	1200 c/s.	120 c/s.	12 c/s.
Aluminium	$a' = 4 \times 10^{-4}$ ($m=1$)	$m \neq 1$	5.4 Mc/s.	540 kc/s.	54 kc/s.	—

It follows that it is impracticable to explore in a mudstone medium with extensive radiators as they would become impossibly unwieldy. But clearly exploration, or flaw detection, in a metal such as aluminium is readily possible at a frequency of a few megacycles per second.

The above relationships appear to agree well with practical results and hold for a wide range of variables. For instance, the frequency of about 30 c/s. can be shown to be about optimum for range of about 5,000 feet in geophysical survey. Where the medium is anomalous, as for example air, the choice of optimum frequency is more difficult but for the rather unfavourable case of air at 90% humidity, frequencies are about one-third of those in Table 2.

One of the most important applications of ultrasonics in the telecommunication and allied fields is the provision of precision timing elements, such as quartz frequency standards, the quartz crystal 'clock' and the provision of loss-free electromechanical filter elements for band-pass networks. It may be noted that in these fields a new substance, ethylene diamine tartrate, seems to promise distinct advantages over quartz because of its higher C_m/C_e ratio, and its loss is quite low—about two or three times that of quartz. Flaw detection or ultrasonic non-destructive testing competes with echo-sounding for leading place among other applications, and several speakers described experience with the British system of flaw detection. In this country attention has been concentrated on pulse systems, but in Germany the c.w. systems of Trost and Pohlmann seem to have had some success and may be superior for certain specialized tasks. The British system scans the body by movement of the transmitting and receiving elements, but consideration is now being given to means for steering the beams to curtail labour in scanning. Other minor improvements include the direct indication of the efficiency of contact between transducer and specimen, a matter of great importance if any quantitative estimate of flaw magnitude is attempted.

Mr. D. O. Sproule described the principle and operation of a commercial flaw detector using the pulse method. The equipment normally operated at $2\frac{1}{2}$ Mc/s., but modified modes of working enabled frequencies down to $\frac{1}{2}$ Mc/s. to be used with the same X-cut quartz crystal probes or transducers. The use of high frequencies permits the radiation to be concentrated in a beam, the half-amplitude semi-angle of the beam at $2\frac{1}{2}$ Mc/s. being about 5° . The beams can be reflected and scattered rather like light beams and, since the wavelength in steel at this frequency is about $2\frac{1}{2}$ mm., flaws of this order of size or even smaller can, in suitable circumstances, be readily detected. The equipment has many points of similarity with radar; a short electrical wave train is converted to mechanical waves and transmitted by one crystal probe, the received waves are picked up by a similar probe and, after amplification and rectification, displayed on an amplitude-range trace. The general schematic diagram of the system and the overlap of the polar diagrams for transmitting and receiving are shown in Figures 3(a) and (b); Figure 3(b) shows how short-range operation can be achieved with special probes in which the crystals transmit through metal wedges. Typical results on a special $1\frac{7}{8}$ in. thick steel test piece are given in Figure 4 (Plate I). In this case the 'flaws' were long drilled holes parallel to the surface and $1\frac{1}{4}$ in. below it, their axes being normal to the line joining the wedge-mounted transducers.

Mr. Sproule said that testing with an oil film was satisfactory on plane surfaces but not with a 'rough-as-rolled' surface. In the latter case some improvement

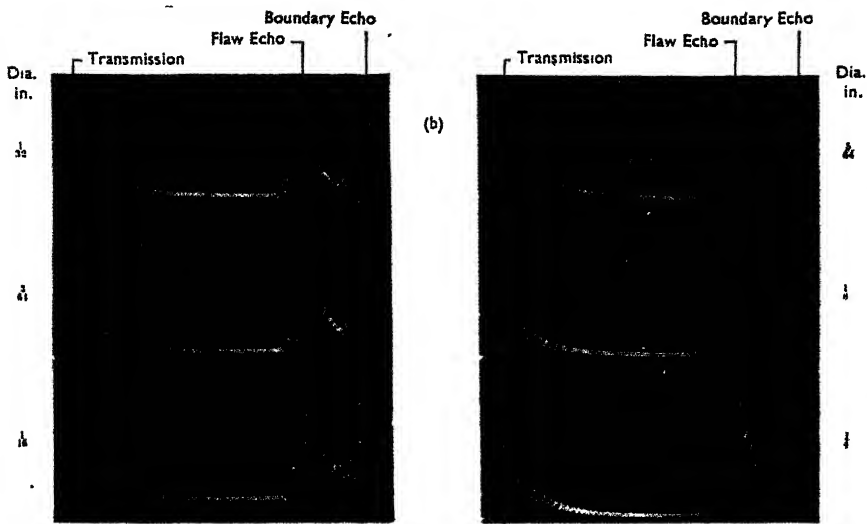


Figure 4. Echoes from cylindrical flaws.

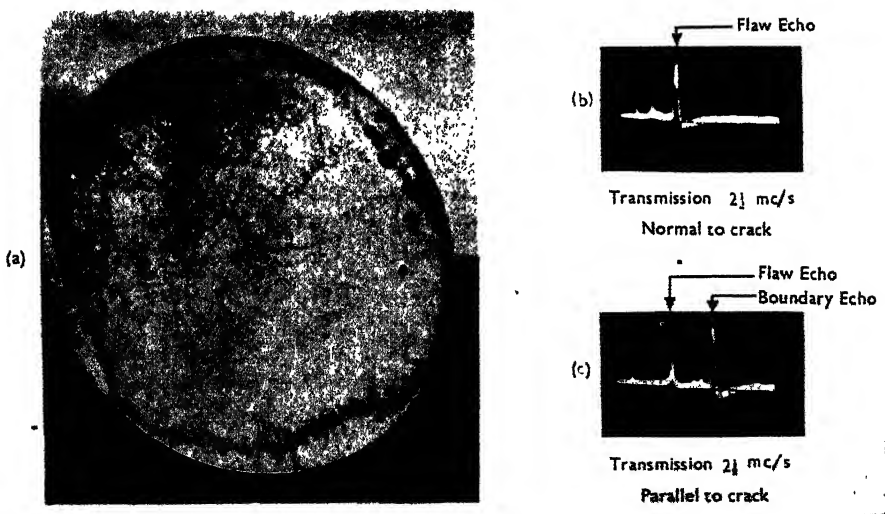
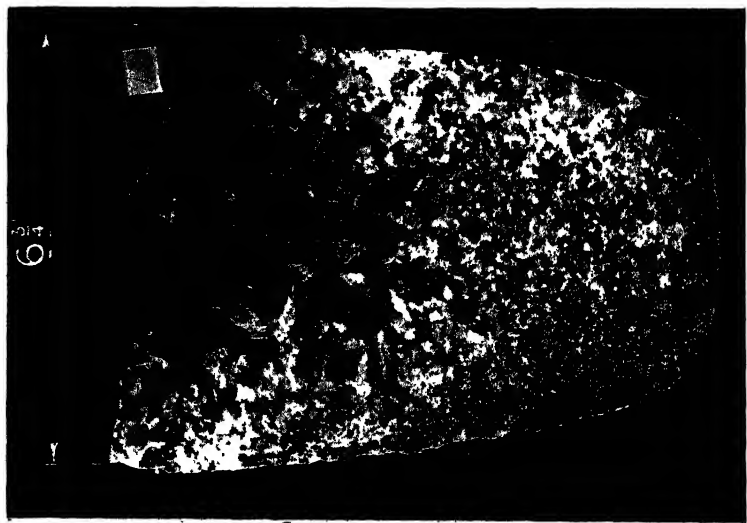
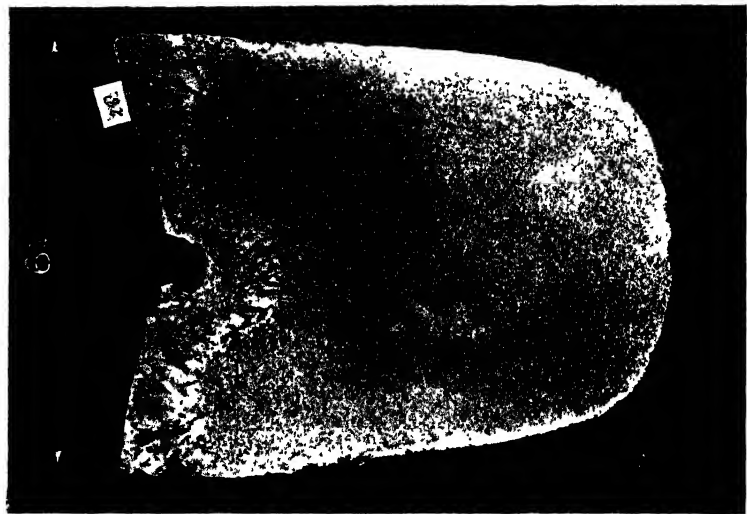


Figure 5. Cast aluminium billet 18 inches dia. showing effect of orientation of large shrinkage crack.



Solidified normally.



Solidified under the influence of ultrasonic agitation.
20 lb. melts of aluminium 4% copper alloy.

resulted from a reduction of frequency from $2\frac{1}{4}$ to $1\frac{1}{2}$ Mc/s., but better results were obtained at $2\frac{1}{4}$ Mc/s. with an amalgam film. Shaped probe surfaces were often helpful.

Rolled and extended materials were generally better than cast, and excellent results were obtained on such materials as aluminium alloys. On porous materials a reduction to $1\frac{1}{2}$ Mc/s. often revealed faults much more distinctly. With care some very good results were achieved: see, for example, Figure 5 which shows even the orientation of a certain flaw in a cast aluminium bar.

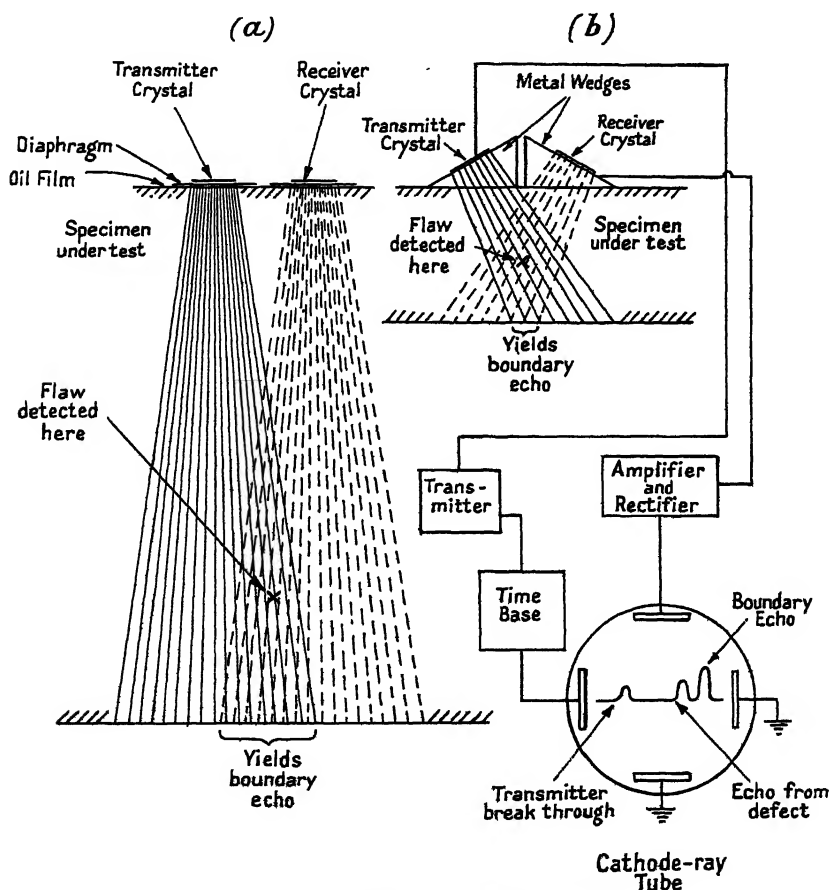


Figure 3. British flaw-detection system.

Mr. G. T. Harris described the use of the flaw detecting equipment for the routine inspection of many thousands of costly austenitic alloy steel forgings for gas turbine rotors which had to withstand very great stress in use at speeds up to 30,000 r.p.m. and temperatures up to 750°C . The rotor blanks were 6 in. thick and 9 in. to 40 in. in diameter, the flat faces being carefully surfaced for application of the probes. This thickness was too great for normal x-ray testing; as the blanks had been flattened, no guidance was obtainable from examination of the surface so that the ultrasonic method seemed without rival. It was found that porosity or defects of the order of half a millimetre or even less could be detected by this means, although it was impossible to differentiate between the presence of a single serious defect and a patch of pinhole porosity or non-metallic

inclusions. However, a grading scheme was successfully adopted for defects and the grading achieved checked consistently on re-examination. Besides examination of the blanks, the original ingots were successfully inspected from machined 45° flats on the corners, the indications of primary and secondary piping and of axial porosity correlating well with subsequent sectioning. The blooms used, which were up to 2 feet diameter, were provided with three or four lengthwise machined flats at 90°, inspection then being carried out by the usual echo method, by observation of the bottom echo or by side-to-side transmission.

Mr. Heselwood described four applications of the above described equipment to the steel industry. The first was to detect 'hairline' cracks, once a serious trouble but now less so owing to improved techniques. The cracks are disc-shaped, about 6 mm. across and $\frac{1}{4}$ micron thick, generally parallel to the surface of the plate, billet or bar. These were readily detected except when few in number and as small as $\frac{1}{2}$ mm. across.

Defects of the lamination type could often be detected with the same equipment but transmitting through from side to side. Quite extensive sheets could be searched with suitable lead arrangements; water with a little wetting agent was used for contact. For plates of sufficient thickness, the reflection method proved more sensitive than the transmission method.

A third application was for testing cold-rolling rolls which were from about 2 in. to 20 in. in diameter and were made by forging to shape followed by a heat treatment. The hardening of the skin was accompanied by high internal stresses often dangerously enhanced by flaws. Ultrasonic inspection has been so successful that no breaks occurred in service after this system was instituted. Finally, an interesting case was described of a 4½ in. square steel ingot which had crystallized with dendrites normal to each face meeting at the diagonal planes. The latter appeared to reflect the rays well while the crystals apparently guided the rays. This structure prevents normal use of ultrasonics in ingots, and such effects together with the 'looseness' often present in steel castings, made flaw detection difficult. Discussing the future, Mr. Heselwood felt that besides improvement (which should be applicable to all types of materials) to facilitate flaw detection close to the surface, there was room for much improvement in the technique for testing iron and steel castings, while a system of rapid scan for large areas would be valuable.

Mr. H. W. Taylor summarized four years' experience in flaw detection on aluminium alloy extrusions and cast billets; routine inspection had been instituted for these cases.

Extrusions required no surface preparation beyond the application of a thin oil film before inspection. Rectangular extrusions could be inspected at the rate of five feet per minute. In the case of extruded bar, which had to be rotated during traversal, the inspection rate dropped to two feet per minute. In rectangular bars he considered that it was possible to deduce the size and configuration of a defect from the flaw detector indications, but correlation was not so good with round bars even with shaped crystal mountings, and the difficulty increased with bars as small as two inches in diameter, although increase in transmitter power was helpful. To avoid misleading results, the probes needed to be in a common tangential plane with the bar. In only one instance was an area of segregate located by ultrasonic inspection.

Cast materials were more difficult to test, and it was often necessary to lower the test frequency. It was generally essential to cut or grind flats on the billets to provide reasonable test surfaces. It was better to test billets from the cylindrical surfaces as effects believed to arise from the orientation of crystals prevented successful testing from the ends. A high-power transmitter was needed when absorption occurred, as in high zinc content alloys, but even these could be tested up to diameters of sixteen inches. Porosity appeared as a mass of reflections varying with small movements of the probes and in such cases a low frequency helped to distinguish defects, although skill was needed to interpret the flaw detector indications. The ultrasonic inspection of cast material was very valuable in detecting gross defects before extrusion, and in monitoring the casting process.

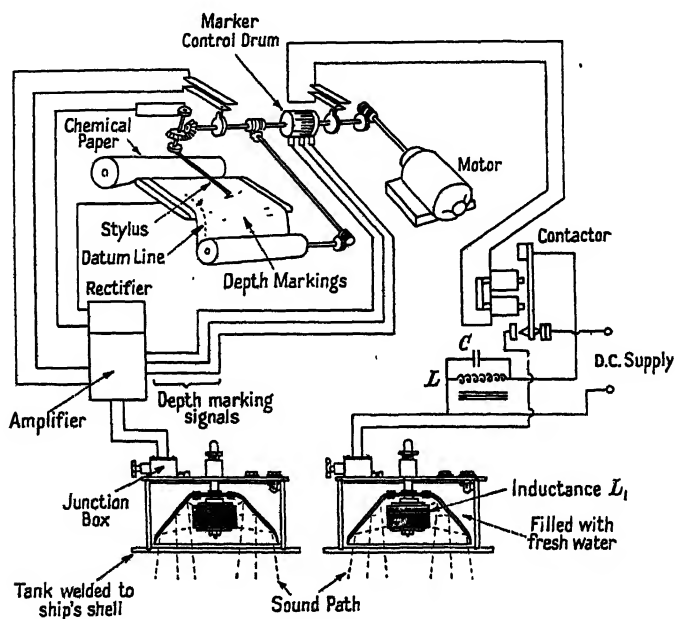
Before describing the next important application, i.e. echo-sounding for ships, mention should be made of the possibilities of this technique in air. Two main applications might be vehicular guidance in fog and an aid to the blind.

In considering the possible use of an ultrasonic guiding device for blind people, it was emphasized that the generation and pick-up of the ultrasonic beam constituted a relatively simple part of the problem, the major difficulties arising from the specular nature of most reflections, the need for extreme lightness and the need to present the picked-up information in a form which could be easily assimilated. Publicity had been avoided for fear of raising the hopes of blind people prematurely. Mr. Beurle said any easy use of ultrasonic devices was unlikely and first described the way in which audible noises were used, surprisingly effectively, to guide blind people. Ultrasonic means might give three advantages, firstly in being inconspicuous, secondly in being less disturbed by ambient noises, and thirdly by increasing the scattering signal from slightly rough surfaces such as walls. A compromise at 50 kc/s. to avoid excessive absorption in air had been found reasonable, and bilamellar or 'bimorph' piezoelectric generating and receiving crystals in parabolic reflectors had worked well. Using a pulse length of about $\frac{1}{2}$ msec. and rate of about 20 p.p.s., it had been found best to combine the rectified transmitted and received pulses after limiting and then to present them aurally to the blind man. Often it was even possible to distinguish between a wall, a fence and a hedge in this way. Although development is not far advanced, the narrow beam ultrasonic device has a strong competitor in a wide-angle device, using binaural direction-finding and operating at the higher audible frequencies. Mr. Beurle felt that, at present, the latter device showed more promise.

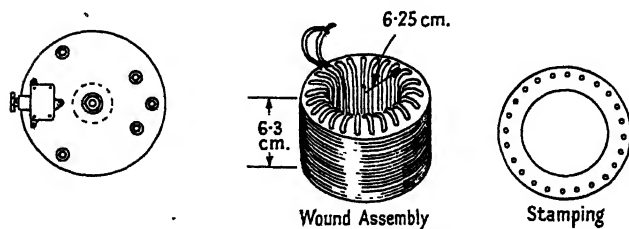
Mr. D. O. Sproule discussed echo-sounding for ships with special reference to a modern magnetostriction type echo-sounding recorder (Figure 6). A toroidal ring of nickel laminations is used as transmitting transducer in the usual pulse system, reception being on a similar transducer. The signal after amplification is marked by a moving stylus on chemical recording paper, the delay in the return of the signal due to its travel time in the water being indicated by the displacement of the mark produced with reference to a datum line. In considering the choice of frequency for transmission, practical considerations may modify somewhat the optimum value deduced from Mr. Bradfield's fundamental equations (8) and (9): high accuracy in range might be required or freedom from the influence of the ship's rolling or of the slope of the sea-bed might be needed, and the water itself might be salt or fresh, aerated or emulsified with mud; the ship's noise spectrum

may also play a part. The equation gives a frequency of 15 kc/s. for 800 metre depth in sea-water, and experience shows this to be about right. A compromise has to be struck as regards the effect of rolling of the ship, since at 15 kc/s., with the equipment sketched in Figure 6, the main lobe semi-angle is about 20° (using 12 to 14 in. aperture), and consequently in heavy weather a record is only obtained for part of the time.

One of the main practical difficulties is the impedance of the skin of the hull which is at least $\frac{3}{8}$ in. thick and in the case of the *Queen Mary* is $1\frac{1}{4}$ in. thick. Furthermore, this thick skin transmits machinery noise which tends to obscure



(a) Schematic diagram of equipment.



(b) Plan view of oscillator mounted in tank. (c) 15 kc/s. magnetostriction transducer.

Figure 6. Echo-sounding recorder of magnetostriction type.

echoes. The high noise level and thick skin reduced range on this ship to 600 fathoms (1,200 yd.) using a frequency of 15 kc/s. and a 5 joule condenser discharge through the magnetostrictive transducer. In some other trials it was found that a range of 1,800 fathoms would be reduced to about 260 fathoms in a storm due to aeration of the water and to noise. As an example of the use of lower frequencies.

for greater range, a 9,000 metre range was obtained at 10 kc/s., using a 10 joule condenser discharge. It is interesting to note that, for short ranges, say 40 feet, even using 15 kc/s., it is possible to measure to an accuracy of 3 inches, which is just under a wavelength in water.

Mr. Sproule concluded by giving details of the equipment employed (see also Wood *et al.* 1935). A motor system drives stylus, paper, marking drum and transmitting and receiving cam-operated contacts roughly as shown in Figure 6 (*a*). When the contactor is operated, the energy in the high-inductance choke L surges into the condenser C , by which time the contactor connects the transmitter to the transducer and a damped train results at $7\frac{1}{2}$ kc/s., the mechanical wave being, however, at 15 kc/s. For a typical discharge of 2 joules, a magnetostrictive transducer as in Figure 6 (*b*) will have a contraction of about 0.004 mm. on the diameter, with radiation about 540 watts at 2 watt/cm², and mechanical Q about 10. The receiver transducer is similar but operated with polarization (the residual flux after saturation). The bandwidth of the amplifier is about $1\frac{1}{2}$ kc/s.; its gain is controlled by an exponentially falling grid bias (initiated by a cam-operated contact) to lessen overloading by the stronger signals occurring at short range.

An ultrasonic application unique in its way is the delay line where times of the order of a millisecond and waves with bandwidths of 1 Mc/s. or more are involved. In television phraseology it was possible to store about 2,000 picture points and change these 500 times a second.

Successful ultrasonic delay lines have been made, using either solid or liquid propagating media, the former being longer, less troublesome in use but more difficult to design. Carrier frequency operation was universally adopted to obtain a uniform phase shift-frequency rate over the wide-frequency band transmitted.

Mr. H. Grayson gave an account of water delay lines pointing out that the $1\frac{1}{2}$ metres required for a millisecond delay corresponded to the travel time in 200 miles of electrical cable. The frequency used, which was always such that the transducer had an aperture of many wavelengths, depended on the bandwidth required. A compromise was necessary between the higher path absorption at the higher frequency, and the higher losses occurring with complex electrical terminating networks when a wide bandwidth with low carrier frequency is essential. Other factors in this choice are the objectionable tilt occurring on the attenuation-frequency characteristic for the high frequencies and the danger of multiple path effects if the attenuation is too low with low frequencies. Using quartz transducers working on one face only, water provides a bandwidth of about 10% with simple terminating networks. Although water has an absorption of about 30 db. at 10 Mc/s. for 1 millisecond delay at normal temperatures, refraction effects have been found troublesome, and it has been found better on this account to work at about 73°C., where the temperature-velocity gradient was zero; fortunately the absorption is also lower there.

Mercury lines were similar in length to water lines; they had the advantage of much greater bandwidth and lower attenuation, but their design and maintenance presented serious difficulties.

Mr. Bradfield said that these accounts did not by any means exhaust applications in this field. He touched on such applications as echo methods in rock and concrete, echo exploration of liquid-filled underground cavities and measurement of velocity and absorption in fluids, hydrosols and aerosols as an aid to analysis.

§3. APPLICATIONS OF THE MECHANICAL FORCES SET UP BY INTENSE RADIATION

Mr. Bradfield then introduced the discussion on high-intensity ultrasonics pointing out that, though spectacular experiments had been made, little industrial application had taken place in this field. In early days effects such as the killing of bacteria had been obscured by the simultaneous heating which occurred, but this complication was now avoided.

In approaching this subject, four points should be borne in mind. (i) Particle sizes did not seem to be critical; for instance, protozoa at say 20μ , bacteria a few microns in size and protein molecules of, say, 100A . all seem vulnerable to a wavelength of 2mm. , i.e. 100 to 200,000 times greater. (ii) There is little evidence of optimum frequencies; indeed effects often persist down to audible frequencies. (iii) The accelerations involved are astonishingly great. (iv) Cavitation, which occurs when the intensity exceeds a few watts/cm^2 , certainly plays a very important part.

Dr. B. E. Noltingk agreed that cavitation was very important in high intensity irradiation, and that nuclei appeared necessary for this to occur so that the effect of impurities would need to be studied closely. As a point of interest, he reported some experimental evidence that cavitation voids may be more like 'hairline' cracks than spherical hollows. Cavitation was not pronounced at a frequency of about 1Mc/s. and oscillatory pressures of about 5 atmospheres but was violent at about 20kc/s. frequency at 3 atmospheres pressure, subsiding into mere bubble formation at 1 atmosphere. Emulsification was especially easy at low frequencies and any pair of immiscible liquids could be emulsified at 20kc/s. (except those having a metal as one constituent).

Mr. G. M. Wells remarked that, in spite of the number of interesting possibilities, there was a dearth of actual industrial application of intense ultrasonic waves. There was a serious lack of precise information as to the way in which actions such as emulsification, dispersion and sterilization depend on frequency. The widespread adoption of the frequency range of 200 to $1,000\text{kc/s.}$ was largely due to the convenience of generation from quartz vibrators. High intensities were readily obtainable in this way, and he had used quartz up to 50watts/cm^2 . Cavitation was clearly of importance and its variation with frequency should be studied. The assumption that its onset coincided with negative pressure in the liquid implied invariance of threshold intensity with frequency, a prediction not borne out by experience, since it had been found easier to produce cavitation at 50kc/s. than at 1Mc/s. Experiments in America by Knapp and Hollander (1948) also showed that the growth of bubbles often lasted some milliseconds, suggesting the periodic time of the wave is important, and the higher intensity with shorter pulses found by Briggs and Johnson (1947) and Mason (1947) provided evidence of rather similar nature. The fact that von Ardenne (1940, 1941, 1943) had found emulsions prepared by violent shaking at 100c/s. to be apparently identical with those prepared ultrasonically showed the wide field of frequency variation which needed to be explored.

Miss M. Thornley described experimental work on bacteriological effects, pointing out first that published work seemed to indicate that oxidation was unimportant (Flosdorf *et al.* 1936, Rouyer and Grabar 1947), but that cavitation increased mortality. It was thought that shock waves were started which were

highly lethal. Such a process might well yield a logarithmic survival curve, and she had indeed found, using 1 Mc/s. radiation up to 2 watts/cm³ (ratio of R.F. input power to sample volume) on *Bacterium coli* in dilute casein solution, that a constant proportion of viable cells was killed per unit time though this proportion sometimes decreased slowly with time. The decrease might be due to some of the organisms being tougher than others. Care was needed to keep temperature constant in spite of heating. Her tests at Imperial College showed evidence at 1 Mc/s. of a threshold, then an increased killing rate as intensity increased; Shropshire (1947), though working at 9 kc/s. and admittedly using a different suspending medium, had found no such threshold. Suspending media played an important part (Beckwith *et al.* 1936) and milk, perhaps on account of the fats it contained, was difficult to sterilize.

On account of the wide variations in conditions reported in published work (Paic *et al.* 1935, v. Euler *et al.* 1944, Chambers *et al.* 1932, Yen *et al.* 1934 and Grabar *et al.* 1945), it was impossible to say whether high or low frequency irradiation was more lethal, though negative results were sometimes reported at high frequencies.

There might be a valuable application in extracting enzymes (Stumpf *et al.* 1946) without damaging them and also in releasing virus from tissue cells (Chambers *et al.* 1936). On the whole, small scale work, especially on valuable materials, seemed most promising, industrial applications having yet to be justified since they might prove uneconomical.

Mr. A. W. Cronshaw reported on bacteriological and emulsifying effects. He used a frequency of 200 kc/s., with 44 watts/cm² intensity at the 2 in. diameter crystal; transformer oil of viscosity 0.2 poise was used. Irradiation of both cultures of *Bacterium coli* and of *Staphylococcus albus* in a test tube held above the crystal for 10 minutes gave no lethal results. He found that emulsification took place far more rapidly at 10 kc/s. than at 200 kc/s.

Mr. C. J. Bradish described his investigations, which were mainly on the irradiation of proteins with ultrasonics. The experiments were carried out at about 700 kc/s. with a power-volume ratio of 6 watts/cm³, very careful precautions being taken to avoid appreciable temperature rise. The first reaction studied was coagulation, and it was found that several proteins had the same coagulation rate of 15% per hour: human serum proteins, human carboxy haemoglobin, egg albumin and *Helix pomatia* haemocyanin. This was a much higher rate than would arise through a simple increase in air-solution interface. The reaction products appeared to be the same as those which would result from coagulation by simple heating; moreover the uncoagulated residue was apparently unchanged. Irradiation was not accompanied by change of gross physical properties, for instance, denaturation which is believed to entail an uncoiling of the constituent polypeptide chain. On the other hand, it has been reported that irradiating normal horse serum for a few hours at 1 to 2 watts/cm³ altered the soluble-fraction.

In contrast to some of the above proteins, with molecular weights around 75,000, was fibrinogen with molecular weight five times as large and length about 0.07 μ . This is an essential agent in the clotting of blood. Irradiation for one hour was sufficient to prolong clotting times tenfold. An even larger molecule, *Helix pomatia* haemocyanin, with a molecular weight of 9×10^6 , length nearly

0.1 micron and diameter one-sixth of this, has the property of longitudinal cleavage to $\frac{1}{2}$, $\frac{1}{4}$ or $\frac{1}{8}$ the size, but Mr. Bradish has been quite unable to confirm Brohult's claim (Brohult 1940) that irradiation caused an increase in the relative concentration of the half molecule. There was, however, a notable polydispersion of the one-eighth molecule due to irradiation.

These reactions have appeared to proceed at a rate proportional to watts/cm², and there was no evidence of a threshold effect. Increased size seemed to be accompanied by increased rate of disintegration; for example, the protozoan *trichomonas foetus* (approximately 15 μ diameter) at 6 watts/cm² suffers 50% mortality in 15 seconds compared with 50% protein coagulation in 3 hours. Mr. Bradish thought that any large unit was subject to considerable frictional forces because it was too bulky to follow the motion of the medium faithfully. Cavitation also plays a part and below $\frac{1}{2}$ micron particle size is probably the most powerful agency in disintegration.

Mr. Maguire described a process which had been successfully used for two years for joining aluminium, especially in the form of fine wires, using 10 kc/s. vibrations, and a power of 50 watts. The vibrator was a hard drawn pure nickel tube 9 in. \times $\frac{7}{8}$ in. o.d. with 1/16 in. wall supported at its centre with the axis vertical. It was slotted at the lower end and was surrounded there with a coil carrying 10 kc/s. and D.C. polarizing currents. The upper end was closed by a brazed steel disc $\frac{1}{8}$ in. thick mounting a 1 in. \times 5/32 in. diameter rod surmounted by a $\frac{1}{4}$ in. \times $\frac{5}{8}$ in. diameter brass cylinder. The latter had a bead of 70% tin 30% zinc solder on top which was kept molten by a gas jet. The intensity of vibration was adjusted to give a striated appearance to the solder. The aluminium, or aluminium and copper wires to be soldered were twisted loosely and, without cleaning, were immersed in the solder globule for some seconds. Wires down to 0.006 in. diameter had been used. There was no subsequent change in the resistance at the joint.

Larger wires and flat aluminium objects could be tinned by being pressed on to the cylinder top, avoiding pitting from too prolonged contact. Subsequently such tinned parts could be joined with ordinary solder at a low temperature with the operation shortened to avoid loss of the special solder surface.

Mr. A. N. Turner discussed the effects which ultrasonic irradiation might cause in metallurgy and described experimental work along three lines involving aluminium alloys. A 120 watt generator of 26 kc/s. vibrations was used and the melts each weighed about 20 lb. The melt at 100° C. superheat was transferred to a refractory lined crucible open at the bottom where it rested on a heavy cast steel base plate. Vibrations were introduced at 50° C. above liquidus and solidification took 15 to 20 minutes. The degassing effect in an aluminium-4% copper alloy was compared with that from degassing with chlorine, the criterion being a test based on the density of specimens taken from the metal after subjection to a special reheating and vacuum cooling treatment; it was found to be equally efficacious.

The second effect investigated was reduction of grain size. Excellent results were obtained on the alloys aluminium-4% copper, aluminium-12% silicon, aluminium-9% silicon and aluminium-0.5% manganese. It was especially noteworthy that the dendrites in the microstructure were broken up by the vibrations and a proposed explanation of the refining action on grain size was that newly formed dendrites in a metastable region during solidification were shattered

by irradiation and so produced showers of new nuclei for the formation of new crystallites. Figure 7 (Plate II) shows a typical irradiated 4% copper alloy compared with a non-irradiated control.

The third effect was the tendency to increase segregation of alloy components. In a silicon alloy, since the specific gravity of aluminium and silicon (at room temperature) are 2.7 and 2.4 respectively, there is a strong tendency for primary silicon in a hyper-eutectic alloy (the eutectic is 12% silicon) to float while there is little tendency for primary aluminium in a hypo-eutectic alloy to sink. Even in the latter alloys, ultrasonic irradiation produces appreciable segregation. This effect is also strongly marked in copper alloys which had been investigated up to the aluminium-33% copper eutectic.

In the course of this work it was felt that an improved method of monitoring the intensity of the vibrations in the melt was greatly to be desired.

Mr. Bradfield summarized by saying that, as a general rule, it was more profitable to use ultrasonic waves for irreversible actions such as killing microbes, cleansing, wetting fine particles, dislodging oxide films and scales, freeing dissolved gases, refining the grain of metals, precipitation of aerosols and drying paper pulp, while for reversible actions such as emulsification and coagulation of hydrosols the method seemed less promising. Care also had to be taken that two effects might not interfere, for instance, segregation of components in irradiating molten metals might take place as well as refinement of the grain. Finally, the effects of standing waves must not be overlooked, and for this and other reasons, it was highly desirable that methods of measuring intensity of irradiation easily and reliably should soon be established and become widely used in investigations.

REFERENCES

- ALLEN, C. H., and RUDNECK, I., 1947, *J. Acoust. Soc. Amer.*, **19**, 847.
 VON ARDENNE, M., 1940, *Colloid Z.*, **93**, 158; 1941, *Angew. Chem.*, **54**, 144; 1943, *Chem. Tech.*, **16**, 177.
 BAUER, E., 1949, *Proc. Phys. Soc. A*, **62**, 141.
 BECKWITH, T. D., and WEAVER, C. E., 1936, *J. Bact.*, **32**, 361.
 BRIGGS, H. B., and JOHNSON, J. B., 1947, *J. Acoust. Soc. Amer.*, **19**, 4, 665.
 BROHULT, S., 1940, *Nova Acta Soc. Sci. Upsal.*, **4**, 12, 4.
 CHAMBERS, L. A., and FLOSDORF, E. W., 1936, *Proc. Soc. exp. Biol. N.Y.*, **34**, 631.
 CHAMBERS, L. A., and GAINES, N., 1932, *J. Cell. Comp. Physiol.*, **1**, 451.
 VON EULER, H., and SKARZYNSKI, B., 1944, *Arch. Kemi. Min. Geol.*, **17 B**, 15.
 FLOSDORF, E. W., and CHAMBERS, L. A., 1935, *J. Immunol.*, **28**, 297.
 FLOSDORF, E. W., CHAMBERS, L. A., and MALISOFF, W. M., 1936, *J. Amer. Chem. Soc.*, **58**, 1069.
 GRABAR, P., and ROUYER, M., 1945, *Ann. de l'Inst. Pasteur*, **71**, 154.
 HALL, L., 1948, *Phys. Rev.*, **73**, 775.
 KNAPP, R. T., and HOLLANDER, A., 1948, *Trans. Amer. Soc. Mech. Engrs.*, **70**, 5, 419.
 KNESER, H. O., 1938, *Ann. Phys., Lpz.*, **32**, 277.
 LAMB, J., and PINKERTON, J. M. M., 1949, *Proc. Roy. Soc. A*, **199**, 114.
 LOISELEUR, J., 1945, *Ann. de l'Inst. Pasteur*, **71**, 378.
 MASON, W. P., 1947, *Trans. Amer. Soc. Mech. Engrs.*, **69**, 359; 1948, *Phys. Rev.*, **74**, 1134.
 MASON, W. P., BAKER, W. O., MCSKIMMIN, H. J., and HEISS, J. H., 1948, *Phys. Rev.*, **73**, 1074.
 PAÏC, M., DEUTSCH, V., and BORCILA, I., 1935, *C.R. Soc. Biol.*, **119**, 1063.
 PINKERTON, J. M. M., 1949, *Proc. Phys. Soc. B*, **62**, 129.

- PRUDHOMME, A. O., and GRABAR, P., 1947, *Bull. Soc. Chim. biol., Paris*, **29**, 122.
 RAPUANO, R. A., 1948, *Phys. Rev.*, **72**, 78.
 ROUYER, M., and GRABAR, P., 1947, *Ann. de l'Inst. Pasteur*, **73**, 215.
 SHROPSHIRE, R. F., 1947, *J. Bact.*, **53**, 685.
 SPAKOWSKI, B., 1938, *C.R. Acad. Sci., U.R.S.S.*, **18**, 169.
 STUMPF, P. K., GREEN, D. E., and SMITH, F. W., 1946, *J. Bact.*, **51**, 487.
 WOOD, A. B., SMITH, F. D., and MCGEACHY, J. A., 1935, *J. Instn. Elect. Engrs.*, **76**, 461, 550.
 YEN, A. H., and LIU, S. C., 1934, *Proc. Soc. exp. Biol., N.Y.*, **31**, 1250.

APPENDIX

List of Authors and Titles of Papers summarized (sequence as in summarized proceedings).

Author	Title of Paper
G. Bradfield	Applications of Ultrasonics.
J. M. M. Pinkerton	On the Study of Fundamental Properties of Liquids and Solids by Ultrasonic Methods.
D. O. Sproule	Flaw Detection in Metals.
G. T. Harris	An Industrial Application of Ultrasonic Testing.
W. C. Heselwood	Flaw Detection in Steel.
H. W. Taylor	Flaw Detection in Light Alloys.
R. L. Beurle	Ultrasonic Guiding Devices for Blind People.
D. O. Sproule	Echo Sounding.
H. Grayson	Ultrasonic Time Delays.
B. E. Noltingk	Cavitation.
G. M. Wells	Use of High Intensity Ultrasonics in Liquids.
Miss M. Thornley	Bacteriological Applications of Ultrasonics.
A. W. Cronshaw	Bacteriological Effects, Preparation of Emulsions.
C. J. Bradish	Ultrasonic Vibrations and the Protein Molecule.
C. R. Maguire	Soldering Aluminium.
A. N. Turner	Achieving a Fine Crystalline Structure during Solidification.

Measurement of the Equivalent Electrical Circuit of a Piezoelectric Crystal

By A. C. LYNCH

Post Office Engineering Research Station, London, N.W. 2

MS. received 17th August 1949, and in amended form 7th November 1949

ABSTRACT. A new method for measuring the reactances in the circuit requires measurements of equivalent capacitance at several frequencies around that of resonance, preferably differing from the resonance frequency by from $\frac{1}{2}$ to 5%; the accuracy is unaffected by capacitance in parallel with that of the specimen. The measurements can be made with a suitably screened substitution Schering bridge; it should be suitable for frequencies between about 50 and 600 kc/s., and the only unusual component needed for it is a variable condenser which can be read to 0.001 pF. If there are no unwanted modes of vibration, the equivalent circuit can be measured to within 0.1%.

The piezoelectric coefficient d is sometimes calculable from the equivalent circuit and the dimensions of the crystal, but there are difficulties which limit the accuracy of d to about $\frac{1}{2}$ %. Results obtained include, for example, d_{11} for quartz at 20° C. = 6.9×10^{-8} cm/k.s.v. of potential.

§ 1. INTRODUCTION

PIEZOELECTRIC effects can be measured with comparable accuracy by either electrostatic or dynamic methods. The latter require observations on bars or plates vibrating in resonance with an alternating electric field. Similar resonance effects are used in the application of crystals in carrier-frequency telephone systems. This paper describes a new dynamic method of obtaining the electrical data which are needed either for deducing the piezoelectric coefficients or in designing and constructing crystal filters for telephone systems.

A bar or plate of piezoelectric material mounted between electrodes, and at a frequency near that of resonance, is equivalent electrically to a circuit containing capacitance, inductance and resistance all in series, shunted by a capacitance approximately equal to the electrostatic capacitance of the specimen (see Figure 1). For two reasons we may need to measure these components.

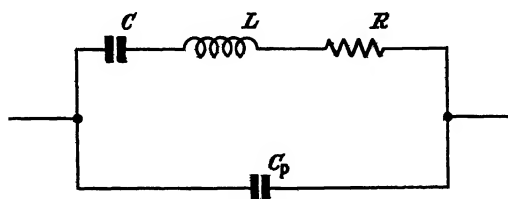


Figure 1. Equivalent circuit of a piezoelectric crystal.

First, when a filter circuit is to include a crystal, data are needed for choosing the orientation and dimensions of the crystal so that it has the required equivalent circuit. Second, the equivalent capacitance C and the resonance frequency f_0 can sometimes be used to deduce the piezoelectric coefficient d (for example, if the specimen is a narrow bar in longitudinal vibration). For specimens of convenient size f_0 will usually be of the order of 100 kc/s. and C of the order of 0.01 pF.

Most of the methods previously used (Cady 1946, p. 387) are based on measurements of both series and shunt resonance frequencies of the crystal. The latter is dependent on the electrostatic capacitance C_p , and on any stray capacitance which may appear in parallel with it. Such stray capacitance is almost unavoidable, particularly if one terminal of the crystal is connected to earth. These methods therefore require an accurate measurement of total parallel capacitance; the present method does not. Few of the published descriptions state the accuracy attainable; for some of the methods it appears to be of the order of within 1% in C . The wide variations between published values of piezoelectric coefficients (see, for example, Cady 1946, p. 220) may be the result either of poor specimens or of errors in the electrical measurements.

The original object of the present work was to obtain an accuracy of within about 1% in C . Since d is derived from C by use of formulae of the type $d^2 = kC/f_0^2$, where k involves the dimensions and density of the crystal and also a numerical constant, this would permit accuracy within $\frac{1}{2}\%$ in d . Later, better accuracy in C was obtained, and this was used to determine the change in C with temperature.

§2. PRINCIPLE OF METHOD

At any one frequency f , the impedance of the crystal is equivalent to that of a capacitance C_x and a conductance G in parallel. At frequencies near that of a resonance of the crystal, the change of the capacitance C_x with frequency is of the type shown in Figure 2. The central part of the curve (especially between the points P and Q) is controlled by the resistance R ; the remainder, which is nearly hyperbolic, is controlled by the capacitance C and inductance L . The

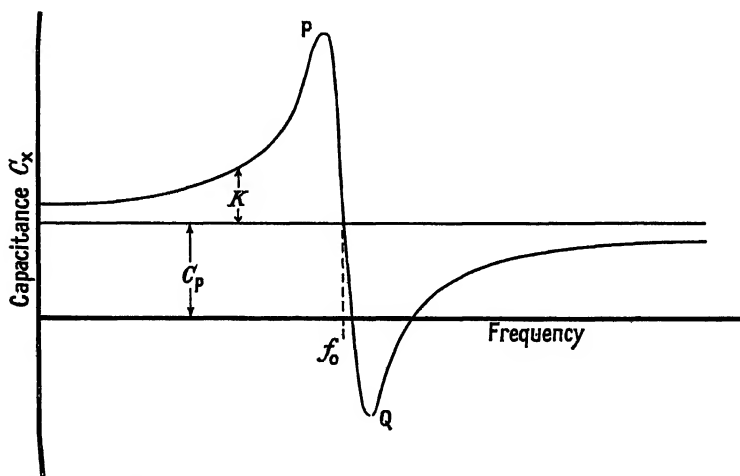


Figure 2. Change of apparent capacitance of a crystal at frequencies near resonance.

curve can, and usually does, cross the horizontal axis, corresponding to negative values of the capacitance C_x . The capacitance difference $(C_x - C_p)$ is introduced by the branch of the equivalent network which includes L , C , and R ; it will be written as K . The capacitance C can be deduced from a group of values of K at suitable frequencies; these frequencies should be in the range for which the curve of Figure 2 is nearly hyperbolic.

If the effect of the resistance R can be neglected, the relation between K and C is simple. At any frequency f , the susceptance of the L -and- C branch of the network is $1/(j\omega L + 1/j\omega C)$, where $\omega = 2\pi f$, and this is equal to the measured susceptance $j\omega K$. Hence

$$K = C/(1 - \omega^2 LC). \quad \dots\dots(1)$$

For two values K_1 , K_2 , observed at frequencies f_1 , f_2 , near the frequency of resonance,

$$\frac{1}{K_1} - \frac{1}{K_2} = \frac{\omega_2^2 LC - \omega_1^2 LC}{C}$$

or, since $\omega_0^2 LC = 1$, where $\omega_0 = 2\pi f_0$ and f_0 is the frequency of series resonance,

$$\frac{1}{K_1} - \frac{1}{K_2} = \frac{\omega_2^2 - \omega_1^2}{\omega_0^2 C} \quad \dots\dots(2)$$

$$\simeq \frac{2\Delta f}{f_0 C} \quad \dots\dots(3)$$

where $\Delta f = f_2 - f_1$, the difference between the two frequencies of measurement. Hence

$$C \simeq \frac{2\Delta f}{f_0(1/K_1 - 1/K_2)}. \quad \dots\dots(4)$$

If in equation (3) we put $f_2 = f_0$, so that $1/K_2 = 0$, we find that the relation between $1/K_1$ and $(f_0 - f_1)$ is nearly linear. Here, then, is a useful method of measuring f_0 : if f_1 and f_2 are on opposite sides of f_0 , we make a linear interpolation between them:

$$f_0 \simeq f_1 + \frac{1/K_1}{1/K_1 - 1/K_2} (f_2 - f_1). \quad \dots\dots(5)$$

Thus, for the present method, the frequencies at which measurements are made can all be integral multiples of a standard frequency (e.g. 1 kc/s.).

If there is more than one resonance frequency, the various equivalent circuits (one for each frequency) can be found separately by successive approximations. By rearranging equation (1), and introducing $\omega_0^2 LC = 1$, we find that the measured capacitance due to a resonance at a frequency greater by F than the frequency of observation is approximately

$$\frac{C'f_0'}{2F(1 + F/f_0')}$$

or, if $F \ll f_0'$, $C'f_0'/2F$ where f_0' is the frequency of that resonance and C' the capacitance in its equivalent circuit. By use of this formula we can correct the observed results for the effect of irrelevant resonances. Graphical methods can also be used.

§3. DISCUSSION OF APPROXIMATIONS

We shall now justify the approximations which have been made; one of them requires the frequencies f_1 , f_2 to be disposed nearly symmetrically about f_0 , and so this condition will be assumed throughout. Put h for the difference between f_0 and the mean of f_1 and f_2 , expressed as a fraction of f_0 ; i.e., $h = [\frac{1}{2}(f_2 + f_1) - f_0]/f_0$. h need not exceed 0.01, even if f_1 and f_2 are restricted to integral numbers of kilocycles per second and f_0 is as low as 50 kc/s. For convenience we shall write $g = \Delta f/f_0$; g may reach 0.2.

There are three approximations to consider:

Approximation (i). We repeat the calculation leading to equation (2), allowing for the effect of resistance R in series with L and C . The measured susceptance must now be equal to the imaginary part of the admittance of L , C , and R in series; it follows that

$$K = \frac{C(1 - \omega^2 LC)}{(1 - \omega^2 LC)^2 + (\omega CR)^2}$$

or, writing a for $(1 - \omega^2 LC)$ and b for ωCR ,

$$K = Ca/(a^2 + b^2).$$

For two values K_1 , K_2 , and corresponding values a_1 , a_2 , b_1 , b_2 ,

$$1/K_1 - 1/K_2 = (a_1 - a_2 + b_1^2/a_1 - b_2^2/a_2)/C,$$

and this can be shown to be equal to

$$(a_1 - a_2)(1 - b_0^2/a_1 a_2)/C,$$

where b_0 is the value of b at frequency f_0 .

But $a_1 a_2 \simeq (g^2 - 4h^2)(1 + h) \simeq g^2$, so that we must consider the likely magnitude of b_0^2/g^2 . Now b is $1/Q$, where Q is the ratio reactance/resistance, and it can be determined either from measurements with the bridge described below or, more conveniently, by a separate experiment. It rarely exceeds 0.001. Hence, for accuracy in C to within 0.1%, this correction should be used if g is less than 0.03. In practice, consistent results are obtained for smaller values of g ; apparently the representation of the crystal by the circuit of Figure 1 is not quite satisfactory when the loss is greater than usual. This is reasonable: for whereas damping by air should be represented by a series resistance as in Figure 1, internal friction or a flaw in the crystal would probably introduce a parallel resistance across part of the inductance and capacitance. Now a parallel resistance, if across the whole circuit, would cause no error in equation (2); so that we expect in practice to find that, in making this correction, b_0 should be given more nearly the value appropriate to an unflawed crystal, which is less than 0.0001. Hence no correction is needed even if g is as small as 0.003.

Approximation (ii). In deriving equation (3) we have put $f_1 + f_2 \simeq 2f_0$. For greater accuracy we could multiply the right-hand side of equations (3) and (4) by $(1 + h)$; but see below.

Approximation (iii). An error may arise in determining the constant capacitance C_p , from which the changes with frequency are measured, because the capacitance of the crystal tends above the resonance frequency towards C_p , but below it towards $C_p + C$. In practice, therefore, we measure K from an initial value between zero and C . Suppose we find this initial value by taking the mean of measurements at frequencies $f_0(1 + m)$, where m is of the order of $\frac{1}{4}$. The mean of these two values of K is $C/(4 - m^2)$, or nearly $\frac{1}{4}C$.

This gives a correcting factor, for the right-hand side of equation (4), of

$$1 - h + \frac{1}{8}g^2h + \frac{1}{64}g^4 - \frac{1}{4}m^2h + \frac{1}{32}m^2g^2,$$

ignoring higher powers than h^2 and g^4 .

Approximations (ii) and (iii) combined. The product of these two factors is

$$1 - h^2 + \frac{1}{8}g^2h + \frac{1}{64}g^4 - \frac{1}{4}m^2h + \frac{1}{32}m^2g^2 \quad \dots\dots(6)$$

and some smaller terms. As m is conveniently about $\frac{1}{4}$, then if $h < 0.01$ and $g < 0.2$ none of the small terms in (6) exceeds 0.00015; and so, for accuracy in C to within 0.1%, no corrections to equation (4) are needed.

§ 4. EXPERIMENTAL APPARATUS

The capacitances can conveniently be measured in a substitution Schering bridge (see, for example, Hague 1943, p. 353). The earth connection should be made to the junction of the resistive arms; thus we exclude from the measured quantity any stray capacitance to earth.

In the apparatus used, the capacitance of the crystal was measured with condensers of either 85 pF. range and 0.01 pF. scale-reading accuracy, or 4 pF. range and 0.001 pF. scale-reading accuracy. These condensers were double-screened. The ratio arms of the bridge were wire-wound, of 100 ohms each. With these values, frequencies from 50 to 650 kc/s. could be used.* The oscillator was of good short-period stability, and its frequency was adjusted to integral multiples of 1 kc/s., or of 250 c/s. if necessary, by comparison with a standard 1-kc/s. supply (accurate to 1 part in 10^7), using stationary patterns on a cathode-ray oscillograph. Balance was observed with a heterodyne detector and telephones. With about 5 v. input to the bridge, balancing to 0.001 pF. was possible. The balance of the bridge, with no specimen connected, was nearly independent of frequency; any variation of the balance could, therefore be regarded as caused by the crystal, and there was no need for balancing both with and without the specimen at every frequency used. In both quartz and ethylene diamine tartrate, the variation of C_p with frequency was found to be negligible, so that, for these materials at least, K is given directly by the change in bridge readings between the test frequency and a frequency remote from that of resonance.

The specimens carried their own electrodes, deposited by evaporation of silver; they were held at their centres between springs carrying silver contacts. The effect of any likely bad centring was negligible (since mounting at two-fifths of the length, a very obvious error, was found to cause an error in equivalent capacitance of only -0.8%). The crystal was screened by a copper box of dimensions 10 cm. \times 10 cm. \times 5 cm. Temperature control was necessary only for measurements of variation of piezoelectric coefficient with temperature, for which the box was heated by a current of warm air. When the ambient humidity was high slight heating was used, to reduce the damping of the specimen.

§ 5. BRIDGE ERRORS

The Schering bridge measures the effective series capacitance of the specimen, not the parallel capacitance which we need. Also, capacitance permanently in parallel with the resistance R_0 of each of the ratio arms causes an error; and capacitance or inductance of the ratio arms themselves will increase or decrease this parallel capacitance. (Possible values of these capacitances would be 200 pF. due to screened leads, transformer windings, and minimum settings of variable condensers, and ± 100 pF. for capacitance or inductance of the ratio arms themselves.) Together these corrections amount to a capacitance

$$C_b = \omega^2 R_0^2 C_2 C_4 \Delta C_4$$

to be subtracted from that observed: C_4 is the capacitance in the arm opposite to the specimen, ΔC_4 the change in C_4 on connecting the specimen and re-balancing, and C_2 the total capacitance in the arm containing the specimen.

But $b \simeq \omega R_0 C_2 \Delta C_4 a^2 / C$; so that $C_b \simeq \omega R_0 C C_4 b / a^2$. Hence for two values a_1, a_2 , such that $a_1 \simeq -a_2$, C_b is nearly constant; its change $\simeq (8h/g^2 - 1)\omega_0 b_0 R_0 C_4 K$, which is normally less than $10^{-5} \times K$ and therefore negligible.

* Even if the ratio arms are not purely resistive, the accuracy of measurement of capacitance in a substitution bridge is almost unaffected.

Bridge settings*				Hence K (pF.)			
Frequency (kc/s.)		In parallel with crystal: C_1 (pF.)		In parallel with ratio arm: C_4 (pF.)		Hence K (pF.) (measured from the mean of settings at 120 and 166 kc/s.)	
134	135	2.754	150	1.017	150	—0.967	
135	136	2.852	150	0.986	150	—0.998	
136	137	2.977	150	1.035	150	—0.949	
137	138	3.147	150	1.127	150	—0.857	
138	139	3.383	150	1.399	149	—0.761	
148	149	0.631	149	2.279	150	mean 1.984	
149	149.5	0.857	149	1.689	150		
149.5		0.950	149				

Bridge settings*				Hence K (pF.)			
Frequency (kc/s.)		In parallel with crystal: C_1 (pF.)		In parallel with ratio arm: C_4 (pF.)		Hence K (pF.) (measured from the mean of settings at 120 and 166 kc/s.)	
149.75	150	1.017	150	1.017	150	—0.967	
150	150.25	0.986	150	0.986	150	—0.998	
151	152	1.035	150	1.035	150	—0.949	
151	152	1.127	150	1.127	150	—0.857	
152	120	1.223	149	1.223	149	—0.761	
120	166	2.279	150	2.279	150	mean 1.984	
166		1.689	150	1.689	150		

Frequency f_1 (kc/s.)	K_1 (from above) (pF.)	Correction for other resonance (pF.)	K_2 (from above) (pF.)	Correction for other resonance (pF.)	Corrected K_2 (pF.)	$\frac{1}{K_2}$ (pF.) ⁻¹	$\frac{1}{K_1} - \frac{1}{K_2}$ (pF.) ⁻¹	$\frac{\Delta f}{\Delta f}$ (kc/s.)	$\frac{1/K_1 - 1/K_2}{\Delta f}$ (pF. kc/s.) ⁻¹
138	1.399		148	—1.353	0.714 ₈	—0.739 ₁	1.454	10	0.1454
139			0.715	—1.454	0.715	—1.454			

B) First approximation for main resonance: 148 $\frac{0.715}{1.454} \cdot 10 = 142.9$ kc/s. and $C \approx \frac{2}{0.1454 \times 142.9} = 0.0963$ pF.

Hence $f_0 \approx 138 + \frac{0.715}{1.454} \cdot 10 = 142.9$ kc/s. and C just found:

C) Approximation for minor resonance, using the values of f_0 and C just found:

149.5	—1.034	1.056	0.022	150.25	—0.949	0.940	—0.009	45	—110	155	0.75	210
149.75	—0.967	1.009	0.042	150	—0.998	0.964	—0.034	24	—29	53	0.25	210

Hence $f'_0 \approx 149.75 + \frac{24}{53} \times 0.25 = 149.86$ kc/s. and $C' \approx \frac{2}{210 \times 149.86} = 0.00006$ pF.

D) Final calculation for main resonance:

134	0.770	152	—0.761	0.002	—0.759	1.29 ₈	—1.31 ₈	2.61 ₇	18	0.145 ₄
135	0.868	151	—0.857	0.004	—0.853	1.15 ₈	—1.17 ₂	2.32 ₄	16	0.145 ₂
136	0.993	150	—0.998	0.034	—0.964	1.00 ₇	—1.03 ₇	2.04 ₄	14	0.146 ₆ †
137	1.163	149	—1.127	—0.005	—1.132	0.859 ₈	—0.883 ₄	1.743 ₂	12	0.145 ₃
138	1.399	148	—1.353	—0.003	—1.356	0.714 ₈	—0.737 ₅	1.452 ₃	10	0.145 ₂

Hence $f_0 = 138 + \frac{0.715}{1.452} \cdot 10 = 142.93$ kc/s. and $C = \frac{2}{0.1453 \times 142.93} = 0.0963$ pF.

Table 2

Type of crystal	Dimensions (cm.)	Frequency at 20° C. (kc/s.)	Capacitance at 20° C. (pr.)	from present work (pF/cm.)	Ct/lb from previously-published work* (pF/cm.)	Piezoelectric coefficients $\times 10^8$ (cm/k.s.u. of potential) in direction given by specimen	Hence d_{11}
X-cut bars	$3.622 \times 0.503_8 \times 0.187_8$	74.86	0.0328 ₈			$\pm 6.9_8$	6.9 ₈
	$3.627 \times 0.504_0 \times 0.199_8$	74.83	0.0296 ₀			$\pm 6.7_9$	
Bars in X-plane :							
45° from Y-axis	$2.001 \times 0.300_8 \times 0.204_8$	138.71	0.0045 ₈			$\pm 4.5_9$	6.9 ₇
135° " "	$2.001 \times 0.300_8 \times 0.204_0$	174.25	0.00194			$\pm 2.3_8$	
CT plates							
-37° 20' from axis	$\left\{ \begin{array}{l} 2.128 \times 2.136 \times 0.053_6 \\ 2.168 \times 2.168 \times 0.054_8 \end{array} \right.$	$\left\{ \begin{array}{l} 145.3_8 \\ 142.9_8 \end{array} \right.$	$\left\{ \begin{array}{l} 0.098_8 \\ 0.100_8 \end{array} \right.$	$\left\{ \begin{array}{l} 0.00116_8 \\ 0.00115_8 \end{array} \right.$	$\left\{ \begin{array}{l} 0.00117_0 \\ 0.00117_0 \end{array} \right.$		
DT plates :							
52° 20' from axis	$\left\{ \begin{array}{l} 2.305 \times 2.305 \times 0.049 \\ 2.306 \times 2.306 \times 0.049_6 \end{array} \right.$	$\left\{ \begin{array}{l} 89.49 \\ 89.45 \end{array} \right.$	$\left\{ \begin{array}{l} 0.116_8 \\ 0.118_1 \end{array} \right.$	$\left\{ \begin{array}{l} 0.00107_8 \\ 0.00109_8 \end{array} \right.$	$\left\{ \begin{array}{l} 0.00108_0 \\ 0.00108_0 \end{array} \right.$		

* Deduced from results given by Bechmann (1943) for plates of similar orientation..

If condensers were used in series with the resistors in the ratio arms instead of in parallel with them, the bridge would measure parallel capacitances and the first of these errors would not arise. But, with any practical condensers, the error due to the initial impurity of the ratio arms would become many times larger.

§ 6. EXAMPLES OF RESULTS

Each of the following examples illustrates both successful use of the method and a difficulty in the measurements. Table 1 gives the results of measurements of a CT-cut quartz plate at 50° c. This plate had two resonances, and the Table shows the method of dealing with the effect of a weak resonance in the presence of a strong one. If there had been a single resonance, the measurements at 149.5, 149.75 and 150.25 kc/s. would not have been needed, and neither would sections B and C of the table, nor the part of section D between the vertical lines.

The results for f_0 and C of a suitable specimen may be used in finding the piezoelectric coefficient for a particular direction—e.g. along the length of a bar (the necessary formulae for narrow bars are well known: see, for example, Cady 1946, pp. 88, 297). These results can then be combined if necessary to give piezoelectric coefficients referred to the customary axes.

Table 2 shows some results for quartz at 20° c. The results for the two X-cut bars did not agree as closely as was expected: the second bar was afterwards found to be partially twinned. The + sign has been used for d_{11} , in agreement with the I.R.E. convention (Cady 1946, p. 409). The coefficients for the 45° and 135° bars must be given – signs, for otherwise they could not be combined to give a reasonable value for d_{11} . The results for d_{11} may be compared with the value 6.9×10^{-8} , recommended by Cady (1946, p. 219) in his review of published results. There are no published formulae for deriving d from the C of plates in contour-shear vibration; the results are therefore compared with those for plates of similar orientations measured by Bechmann (1943) by a different method. The results for both bars and plates are in good agreement with the earlier work.

§ 7. CONCLUSION

The consistency obtained in measurement of the equivalent capacitance C is usually of the order of within 0.1%, provided that C is at least 0.02 pF. and that there is no other resonance close to that being studied. The method has been used to measure C 's of from 0.0002 to 0.7 pF., at frequencies ranging from 50 to 650 kc/s.

On the evidence of Table 2, and similar groups of results for other materials, both at 20° c. and at 50° c., it seems reasonable to claim an accuracy reaching to within about $\frac{1}{2}\%$ in the piezoelectric coefficient. With an idealized specimen the error in d would be only half that in C . The additional error found in practice is caused by departure of the experimental conditions from those assumed in the formulae giving C in terms of dimensions etc. There may, for example, be imperfections in the shape and mounting of the crystals, multiple resonances, or, in quartz, twinning. When measurements are made at two temperatures the effect of these imperfections is usually fairly constant, and, if so, the temperature coefficient of d can be found with the accuracy available in measurements of C , or even better. Some results obtained by this method have already been published (Bechmann and Lynch 1949) and others are awaiting publication.

ACKNOWLEDGMENTS

Acknowledgment is made to the Engineer-in-Chief of the General Post Office for permission to make use of the information contained in this paper. The author is indebted to members of the Radio Branch of the Post Office Engineering Department for providing many accurately made specimens.

REFERENCES

- BECHMANN, R., 1943, *Hochfrequenztech. u. Elektroakust.*, **61**, 1.
BECHMANN, R., and LYNCH, A. C., 1949, *Nature, Lond.*, **163**, 915.
CADY, W. G., 1946, *Piezoelectricity* (New York: McGraw-Hill).
HAGUE, B., 1943, *Alternating-current Bridge Methods* (5th edition) (London: Pitman).

Surface Effects and the Plasticity of Zinc Crystals

BY S. HARPER AND A. H. COTTRELL

Metallurgy Department, University of Birmingham

MS. received 16th October 1949

ABSTRACT. The effects of various surface treatments on the plastic properties of zinc crystals are examined. Roscoe's observation of the hardening effect of a surface film of oxide is confirmed; electrolytically polished specimens have critical shear stresses of about 33 gm.mm^{-2} whereas oxidized specimens have values up to 65 gm.mm^{-2} . Similar effects are observed on the rate of creep under constant stress. The Rehbinder effect—softening crystals by immersing them in paraffin and oleic acid—can be produced on oxidized specimens but not on others. Oxidized specimens do not respond instantly to immersion and the time of response is proportional to the viscosity of the liquid. It is suggested that the liquid penetrates the oxide film and reduces its ability to harden the metal; the action of the oxide film itself is not yet clear.

§ 1. INTRODUCTION

THE success of Griffith's theory of cracks in explaining the low breaking strength of brittle solids has encouraged occasional attempts to explain the low yield strength of plastic crystals in a similar manner, although interest has mostly been diverted in recent years to the theory of dislocations. One result of efforts to obtain evidence for the action of cracks has been the observation that the plastic properties of some metal crystals are affected by the surface condition of the specimens. Roscoe (1936) showed that oxide films of about 1,000 atoms thickness on cadmium crystals raised the critical shear stress of these to 2.4 times that of freshly cleaned crystals. This has been confirmed recently by Cottrell and Gibbons (1948). In explanation Roscoe suggested that molecules of oxide sealed up surface cracks and strengthened the outer layers of the crystal. Using crystals of zinc and tin, Rehbinder, Lichtmann and Maslenikov (1941) observed that the yield strength was the same for specimens tested in air and in a non-polar paraffin oil, but that, when 0.2% of oleic acid was added to the oil, the strength was halved. Similarly, the rate of flow in a creep test could be increased five to ten times by the addition of oleic acid. These effects were accompanied by an increase in the number of glide lamellae and a decrease in their size. It was suggested that the oleic acid acted by penetrating surface cracks, causing them to open and expand.

Although these observations show that plastic properties are affected by surface conditions they do not go far towards indicating the mechanism involved. It was felt that further investigations were needed and the work on zinc crystals reported below was started. While this was in progress the results of two other investigations became known. Andrade and Randall (1948) reported experiments in which clean cadmium crystals were heated in air or *in vacuo*, or were immersed in solutions of cadmium and other salts. Heating in air for 20 hours at 200° c. produced hardening and this was attributed to the formation of a surface film, in agreement with Roscoe's work. This hardening was produced more rapidly when the specimen was first contaminated locally with cadmium nitrate solution. In general, immersion in salt solutions caused the specimen to soften, although in the case of cadmium nitrate this was followed by a considerable hardening effect. Kemsley (1949) has recently attempted to reproduce the Rehbinder effect on tin crystals; although the technique used in the original work was followed he was unable to obtain an observable response to treatment in oleic acid solution.

§ 2. EXPERIMENTAL DETAILS

All experiments were made on a sample of spectroscopically pure zinc supplied by Messrs. Johnson Matthey and Co., in the form of 1 mm. wire. A modification of the Kapitza method due to Andrade and Roscoe (1937) was used to grow crystals. By cleaning each wire before growth, using baked silica quills to support the wires while molten and an atmosphere of argon to reduce oxidation and volatilization, crystals could be prepared which had smooth, clean surfaces and which could be removed easily from their quills. Each crystal was cut into three pieces with the aid of a fine gas flame, care being taken to localize the heating and to minimize distortion. Two of the pieces, each 4 cm. long, were used for mechanical tests and the remaining one for determining the crystal orientation. Barrett and Levenson's (1940) etch-pit method was used for orientation measurements, with an accuracy within about 1°. Some of the specimens for mechanical testing were electropolished, using the chromic acid solution suggested by Rodda (1943). These were each immersed in the solution along the axis of a nickel tube, which served as the cathode and an E.M.F. of 6 volts was applied for 10 seconds. This gave a highly polished surface free from irregularities and reduced the diameter by about 0.1 mm.

In the early part of the investigation, the effect of the surface condition on the critical shear stress was studied. Most of these experiments were made on an apparatus of the type described by Andrade and Roscoe (1937). The load-extension curve was recorded photographically and it was possible to extend specimens at fairly fast rates, about 10^{-2} cm.sec⁻¹. A few experiments were also made at lower rates of extension, 10^{-4} – 10^{-5} cm.sec⁻¹, with an apparatus similar to that described by Polanyi (1925). As is usual in critical shear stress measurements a large scatter was observed, amounting occasionally in extreme cases to 20% of the average value. However, more reproducible results were obtained among specimens taken from the same crystal than from different crystals, suggesting that the scatter was due to some uncontrolled variation, for example in the gas content, among the crystals. To reduce the effect of the scatter repeated tests on several specimens were always made and their results averaged.

Later experiments were made by creep testing, using an apparatus which enabled constant shear stress to be maintained on the glide plane in the glide-

direction (Cottrell and Aytakin 1947, 1950). Creep testing had the advantage that the surface condition of a specimen could be changed while the test was in progress, so that difficulties due to variations in the properties of different specimens were avoided.

§ 3. COMPARISON OF VARIOUS SURFACE CONDITIONS.

Table 1 summarizes the measurements of the critical shear stress after applying various surface treatments.

Untreated specimens (Series A) possessed light-grey oxidized surfaces and an average critical shear stress of 56 gm.mm^{-2} . Series B and C represent various treatments, the common feature of which was a light attack by dilute hydrochloric acid. Apart from the C5 specimens, a critical shear stress of about 43 gm.mm^{-2} was observed in all cases, independently of whether the specimen was tested while immersed in the acid, made the electrode in a cell, or removed from the acid

Table 1. Critical Shear Stresses of Zinc after various Surface Treatments
(Rate of Strain $\approx 10^{-2} \text{ sec}^{-1}$)

Serial number	Surface treatment	Number of tests	Average stress for 0.5% extension gm.mm^{-2}
A	Freshly grown, untreated	15	56
B	1 Tested in 10% HCl solution	10	42
	2 As B1, with specimen as anode of electrolytic cell	6	42
	3 As B1, with specimen as cathode	6	42
C	1 Etched in HCl, washed, dried, exposed to air for 5 minutes	6	46
	2 As C1, exposed for 30 minutes	7	42
	3 As C1, exposed for 1 hour	4	45
	4 As C1, exposed for 18 hours	7	44
	5 As C1, exposed for several days	8	56
D	1 Electrolytically polished, washed, dried, exposed to air for 5 minutes	8	33
	2 As D1, exposed for 1 hour	12	33
	3 As D1, exposed for 18 hours	6	33
	4 As D1, exposed for several days	3	32
E	1 Polished, immersed in FeSO_4 solution, dried, exposed to air for 5 minutes	6	41
	2 As E1, exposed for 1 hour	12	46
	3 Polished, immersed in HCl, dried, exposed to air for 5 minutes	6	46
	4 As E3, exposed for 30 minutes	12	44
	5 Polished, dried without washing, exposed to air for 5 minutes	6	41
	6 Polished, immersed in ZnCl_2 solution, dried, exposed to air for 5 minutes	6	44
F	1 Polished, etched in HCl, steamed for 1 hour	6	65
	2 Polished, washed, dried, steamed for 1 hour	6	33
	3 Polished, etched in HCl, steamed for 1 hour, repolished	3	36

and washed and dried before testing. On all the specimens the acid removed the original film and gave the surface a bright matt appearance. The specimens in C5, on the other hand, were exposed to the atmosphere for several days, after etching, until their surfaces resembled those of untreated specimens.

The effect of electrolytic polishing is shown by series D in which specimens were polished, washed, dried and exposed to the atmosphere for various periods of time. Polishing produced brilliant, mirror-like surfaces which resisted atmospheric attack remarkably well, remaining unchanged in appearance for several weeks. This is reflected by the critical shear stresses which were about 33 gm.mm^{-2} in all cases.

Series E summarizes attempts to increase the critical shear stress of polished specimens by the subsequent application of aqueous solutions of various agents, including the chromic acid solution used in the polishing process (E5). The critical shear stress in all cases was, within experimental error, the same as for those specimens treated only in hydrochloric acid (Series B and C). Thus the value of 43 gm.mm^{-2} appears to be characteristic of the lightly attacked condition, independently of the details of the treatment.

Finally, the specimens in Series F were exposed to steam with the object of producing thick oxide coats. Thick, grey coats were formed when etched specimens were treated in steam and the critical shear stress was raised to 65 gm.mm^{-2} ; in the case of polished specimens the treatment had no effect on either the surface appearance or the critical shear stress, again demonstrating the high resistance of the polished surface.

The last result, F3, shows that the high strength produced by treating an etched specimen in steam does not survive a subsequent polishing treatment, and hence that it is caused by the surface state of the crystal rather than by any internal metallurgical changes which might conceivably be produced.

§ 4. SURFACE EFFECTS AT THE START OF PLASTIC FLOW

The above experiments show that the hardening effect of a surface film occurs also with zinc, as well as with cadmium, crystals. In explanation, one might suggest that molecules of oxide or other corrosion products fill surface cracks and thereby reduce the ability of these to concentrate stress. In this case the effect ought to be most marked at the start of a mechanical test, before enough plastic flow has occurred to smooth out stress concentrations. Roscoe (1936) showed that the increased strength of an oxidized specimen persists for large deformations (300% extension), which makes this explanation doubtful, but did not examine the very early stages of plastic flow in detail.

Accordingly, a study was made of the early part of the stress-strain curve, critical shear stress measurements being taken at extensions of 0.05, 0.1, 0.5 and 1%. Electrolytically polished and lightly etched specimens were studied and, to ensure a reliable comparison, the specimens were selected so that, of each pair taken from the same crystal, one was etched and the other polished. Six crystals were examined, using a rate of strain of about 10^{-2} sec^{-1} , and the sets of critical shear stress values for the two surface conditions were averaged to give the results in Table 2. These show that the critical shear stresses start at about the same value and that the greater strength of the etched specimens does not appear until some plastic deformation has occurred. These experiments were repeated at a lower rate of strain, 10^{-5} sec^{-1} , and the results of these in Table 2

show that the difference between the polished and etched specimens is no longer detectable at this rate of strain. Considering also the first result, it thus appears

Table 2. Critical Shear Stresses of Polished and Etched Zinc Crystals at various Extensions

Approximate rate of strain sec ⁻¹	Surface treatment	Average stress, gm.mm ⁻² , for an extension of			
		0.05%	0.1%	0.5%	1.0%
10 ⁻²	Etched	30	32	40	44
	Polished	28	29	33	36
10 ⁻⁵	Etched	21	22	24	25
	Polished	21	22	24	25

that the greater strength of the etched specimens is a dynamical effect connected with the amount and rate of plastic flow, rather than with the stress required to start flow.

§ 5. CREEP EXPERIMENTS

A few creep experiments were made at room temperature to confirm the critical shear stress results. It has been shown (Cottrell and Aytakin 1947, 1950) that the flow of zinc crystals, under constant shear stress on the glide planes in the glide direction, obeys the equation

$$\gamma = \gamma_0 + \beta t^{1/3} + \kappa t,$$

where γ is the shear strain, t is the time and β and κ are coefficients. This equation is essentially the same as that proposed by Andrade (1910) for the creep of polycrystalline metals. In the present work fairly short durations of loading (2-3 hours) were used so that most of the flow was of the transient, β -flow, type. When a specimen had settled down after loading to a well-defined transient flow, its surface condition was changed by immersing it in a suitable fluid, and the effect on the creep curve was studied. The apparatus was arranged so that the fluid could be brought round the specimen without interrupting the creep test or disturbing the specimen.

Figure 1 shows the strain-time curve of a specimen which was heavily oxidized in steam and then subjected to a resolved shear stress of 70 gm.mm⁻². After flow had occurred for 37 minutes the specimen was immersed in dilute hydrochloric acid. The rate of flow immediately increased; a large instantaneous strain took place and was followed by a fast transient flow. The large reduction in the critical shear stress caused by removing the oxide film is thus reflected by the creep behaviour.

The results of a similar experiment on a polished specimen stressed to 39 gm.mm⁻² are shown in Figure 2. Immersion in hydrochloric acid, after 40 minutes under load, produced a perceptible but small decrease in the rate of flow. Since the latter is a much more sensitive property than the critical shear stress, it is thus hardly surprising that the measurements summarized in Table 2 showed no difference between the polished and etched specimens at low rates of strain. A third creep specimen was etched in hydrochloric acid, washed, dried and then made to flow under a stress of 42 gm.mm⁻². Hydrochloric acid was introduced after 30 minutes, but in this case it caused no observable change in the rate of flow.

§ 6. THE REHBINDER EFFECT

An examination of the effects of paraffin and oleic acid on the creep behaviour of specimens given various surface treatments proved interesting. Polished or etched crystals were made to flow under constant stress as before and then, at a suitable stage in the test, they were immersed in B.P. liquid paraffin either with or without 0.2% oleic acid in solution. In all cases, no change in the creep curve could be detected; Figure 3 shows a typical example with a polished specimen

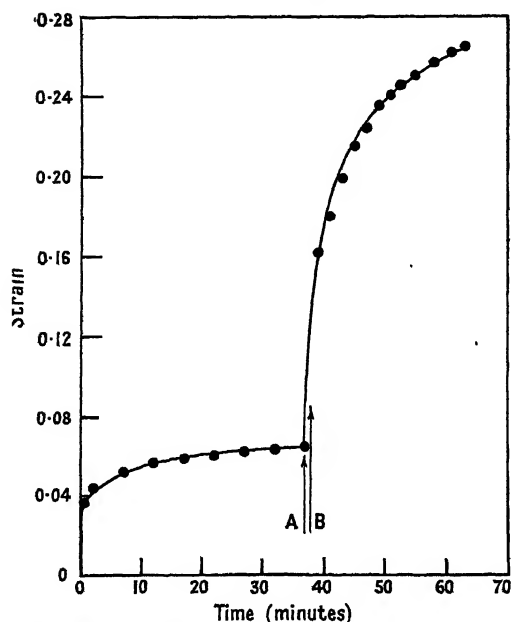


Figure 1. Creep curve of an oxidized crystal. Hydrochloric acid was applied at A and removed at B.

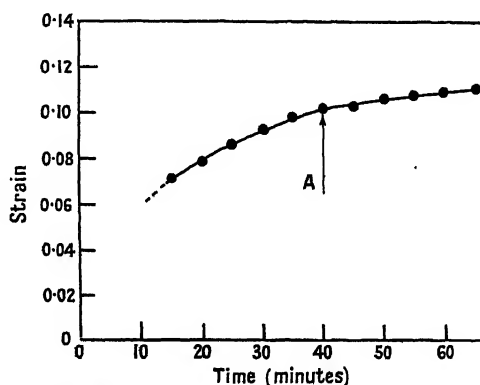


Figure 2. Creep curve of a polished crystal. Hydrochloric acid was applied at A.

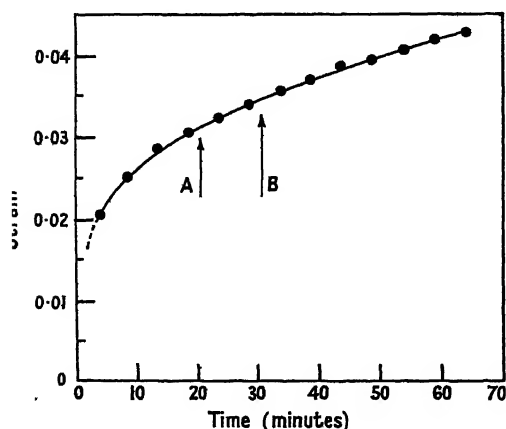


Figure 3. Creep curve of a polished crystal. Paraffin was applied at A and oleic acid solution at B.

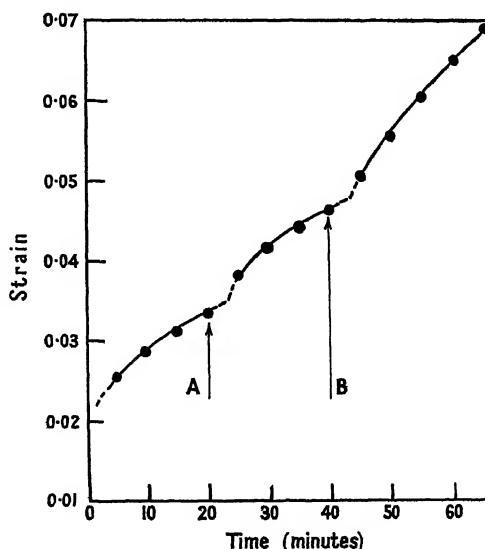


Figure 4. Creep curve of an oxidized crystal. Paraffin was applied at A and oleic acid solution at B.

stressed to 32 gm.mm^{-2} . On the other hand, specimens that had been heavily oxidized in steam responded to the paraffin treatment. Figure 4 shows the creep curve of one such specimen cut from the same crystal as that of Figure 3 under a stress of 63.5 gm.mm^{-2} . At the point A paraffin was introduced round the specimen and, within five minutes of immersion, the rate of flow increased. A second increase occurred when the paraffin was replaced by the solution of oleic acid (point B). Paraffin by itself produced a response only in heavily oxidized specimens and with lightly oxidized ones it was necessary that oleic acid should be present.

It was observed that the creep rate of heavily oxidized specimens could be increased by immersing them in kerosene or ethyl alcohol, whereas polished or etched specimens showed no response to these agents. Also, the creep rate of an oxidized specimen increased almost immediately when the specimen was immersed in kerosene or alcohol; when paraffin was used, on the other hand, a period of about five minutes elapsed before the specimen responded.

§ 7. RELATION OF TIME OF RESPONSE TO VISCOSITY OF MEDIUM

The facts that the Reh binder effect could be produced only on specimens with thick oxide coats and that, with a viscous agent such as paraffin, it appeared some five minutes after immersion, while mobile agents such as kerosene or alcohol produced an almost immediate response, suggested that the effect was concerned with the penetration of the surface film by the agent. This would mean that the time for the response to appear should be related to the viscosity of the agent.

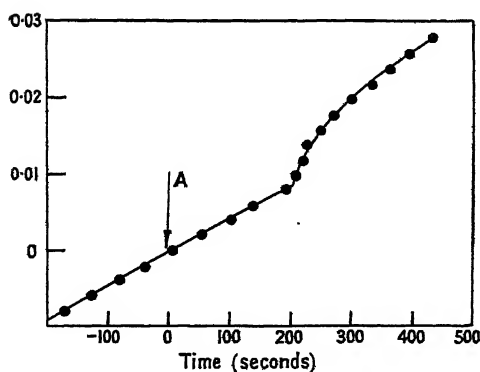


Figure 5. Creep curve of an oxidized crystal. A paraffin-kerosene mixture was applied at A.

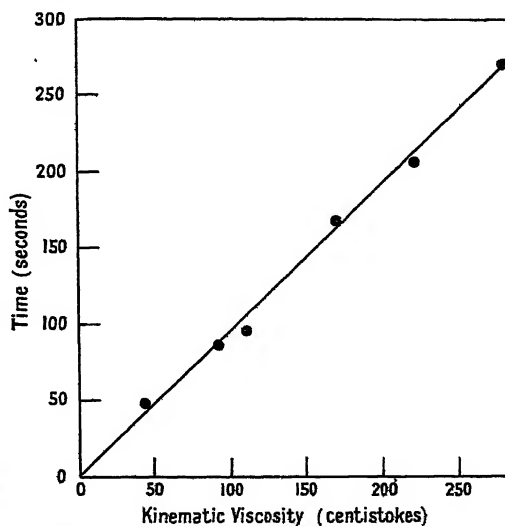


Figure 6. Variation of time of response with viscosity of medium.

Some experiments were made to examine this point using specimens that were standardized, as far as possible, by oxidizing them together in steam. A series of liquids of different viscosities was prepared by mixing paraffin and kerosene in various proportions and their kinematic viscosities were measured at room temperature with the aid of standard U-tube viscometers. Creep experiments were then made and the time interval between immersion in each liquid and the subsequent increase in the creep rate was determined. Figure 5 shows the

response of a specimen to immersion, at the point A, in a liquid of high viscosity (220 centistokes); the creep rate remained unaltered for about 200 seconds and then quite suddenly increased. The results of these experiments are given in Figure 6 which shows that the experimental points lie reasonably well on a straight line through the origin, and hence that the time of response varies linearly with the viscosity.

§ 8. DISCUSSION

The above results provide a possible basis for correlating the observations of Roscoe, Rehbindler and Kemsley. It appears that the primary surface effect is the hardening produced by a film of oxide, or of other products of corrosion, and that paraffin and oleic acid produce a secondary effect by modifying the action of the film. If the surface film is sufficiently thin to be fairly ineffective, on its own account, the oleic acid solution produces no response; this may be the reason why Kemsley was unable to produce the Rehbindler effect.

The response of oxidized specimens to paraffin or similar liquids is not instantaneous and the time of response increases linearly with the viscosity of the liquid. It is thus reasonable to suppose that these liquids act by penetrating the oxide film, presumably through cracks, and weakening its adherence to the surface of the metal, thereby reducing its hardening effect. The mechanism by which the film hardens the metal remains obscure. It is difficult to reconcile a simple explanation, in terms of the ability of surface cracks to concentrate stress, with the observation that the effect appears to be related more to the progress than to the start of plastic flow. Roscoe's suggestion that the outer layers of the metal become hardened by included filaments of oxide is more reasonable.

The recent theory of Frank (1948) concerning the formation of slip-bands by the multiplication of fast dislocations suggests that the surface condition may affect the development of slip-bands and, therefore, the coefficient of strain hardening. This offers the possibility of an alternative explanation, but more work is needed before the effect of the oxide film can be properly understood.

ACKNOWLEDGMENTS

This research was carried out in the Metallurgy Department of the University of Birmingham under the general supervision of Professor D. Hanson, to whom the authors' thanks are due for interest and help. The work forms part of a programme supported by the Department of Scientific and Industrial Research through the National Physical Laboratory, to whom grateful acknowledgment is made.

REFERENCES

- ANDRADE, E. N. DA C., 1910, *Proc. Roy. Soc. A*, **84**, 1.
 ANDRADE, E. N. DA C., and RANDALL, R. F. Y., 1948, *Nature, Lond.*, **162**, 890.
 ANDRADE, E. N. DA C., and ROSCOE, R., 1937, *Proc. Phys. Soc.*, **49**, 152.
 BARRETT, C. S., and LEVENSON, L. H., 1940, *Trans. Amer. Inst. Min. Met. Engrs.*, **137**, 76, 112.
 COTTRELL, A. H., and AYTEKIN, V., 1947, *Nature, Lond.*, **160**, 328; 1950, *J. Inst. Met.*, in press.
 COTTRELL, A. H., and GIBBONS, D. F., 1948, *Nature, Lond.*, **162**, 488.
 FRANK, F. C., 1948, *Report on the Strength of Solids* (London: Physical Society), p. 46.
 KEMSLEY, D. S.; 1949, *Nature, Lond.*, **163**, 404.
 POLANYI, M., 1925, *Z. tech. Phys.*, **6**, 121.
 REHBINDER, P., LICHTMANN, V. I., and MASLENIKOV, V. M., 1941, *C.R. Acad. Sci., U.R.S.S.*, **32** (2), 125.
 RODDA, J. L., 1943, *Min. and Met.*, **43**, 323.
 ROSCOE, R., 1936, *Phil. Mag.*, **21**, 399.

Yield Points in Zinc Crystals

BY H. L. WAIN AND A. H. COTTRELL

Metallurgy Department, University of Birmingham

MS. received 16th October 1949

ABSTRACT. It is shown that sharp yield points can be produced in crystals of zinc containing nitrogen. The gas is introduced in a preliminary melting operation, either by bubbling it through the metal or by allowing it to be absorbed from the atmosphere, using a flux to keep the surface clean. Strain ageing treatments are needed to develop the yield point clearly, but once it is developed it returns at the same level after each successive treatment at the same temperature. Ageing experiments at room temperature show that overstrained zinc crystals soften by recovery before the yield point returns; zinc differs from iron in this respect. General conditions for producing yield points in metals of common crystal structures are briefly discussed.

§ 1. INTRODUCTION

THE phenomenon of the sharp yield point, familiar in mechanical tests on mild steel, is now recognized as a property of some metals containing certain other elements. It appears prominently in iron if the metal contains small amounts of carbon or nitrogen, but not otherwise (Edwards, Phillips and Jones 1940, Snoek 1941, Low and Gensamer 1944). It has also been observed in single crystals of cadmium (Smith 1947, Cottrell and Gibbons 1948) and in polycrystals of molybdenum (Túry and Krausz 1936, 1937), and in both cases nitrogen has been shown to be effective in producing the yield point. A recent theory (Cottrell 1948, Nabarro 1948, Cottrell and Bilby 1949) suggests that the yield point is caused by the segregation of solute atoms to dislocations. The attraction of the dislocations to the segregated atoms provides a bond which has to be broken by a larger force than is necessary to maintain freed dislocations in motion; the material thus gives way suddenly and softens, at the start of plastic flow, producing a sharp yield point. The theory also explains the observed removal of the yield point by plastic overstrain and its return on strain ageing. A freshly strained specimen contains freed dislocations and does not show a yield point, but on ageing these dislocations become anchored by the migration of solute atoms to them and the yield point returns.

Some time ago, Orowan (1934, 1940) observed a yield point of a very similar kind to that described above in single crystals of zinc. Because the effect appeared in crystals that were annealed after overstraining, he called it 'thermal hardening'. However, it seems preferable to retain the terms 'yield point' and 'strain ageing' since these are familiar and adequate, and since the term 'thermal hardening' has also been applied to another effect, the hardening produced by surface oxidation (Andrade and Randall 1948).

Orowan's observation has so far remained unconfirmed. Accordingly, the attempt described below was made to produce yield points in zinc crystals and to ascertain whether the effect required the presence of a specific additional element.

§ 2. METHOD

Two grades of zinc were used, the first being a sample of spectrographic purity kindly presented by the National Smelting Company, and the second a sample of commercially pure zinc. The spectrographic analyses of 1 mm. diameter wires, extruded from these materials, are given in the Table.

Analyses of Zinc Wires

Sample	% Pb	% Cd	% Cu	% Fe
Pure zinc	0.001	0.0001	0.0005	0.001
Commercial zinc	0.010	0.010	0.003	0.003

Single crystals were grown from the wires by Andrade and Roscoe's (1937) fusion method and their orientations were determined by optical measurements of the reflections from etch-pits (Barrett and Levenson 1940). Mechanical tests were made with a machine of the Polanyi (1925) type, using incremental loading, in which 3 cm. wire specimens were strained in tension by means of hooks which engaged with loops of copper wire welded to the ends of the specimens. A standard strain ageing treatment was always given before testing for the presence of a yield point. The specimen was extended a small amount ($\sim 0.3\%$) and, while still mounted in the machine, was annealed in a paraffin bath for 40 minutes at 180°C ., after which it was cooled to room temperature. The test for the yield point was made before the bath was removed, in order not to disturb the specimen between annealing and testing, because of the possibility that mechanical shocks might overstrain the material locally and thereby prevent the observation of a yield point.

The straining, ageing and testing cycle was applied several times to each specimen. The absence of a yield point in the first test or so was considered to be inconclusive since it was not possible to be sure that the crystal was sufficiently straight, that the load was applied axially or that the concentrations of stress near the ends of the specimen were negligible. A high localized stress can cause yielding to take place prematurely, before the mean stress has reached the level of the lower yield point, in which case a smooth stress-strain curve may be observed. However, after a few repetitions of the testing cycle, inhomogeneities in the stress are usually smoothed out by plastic flow and the yield point will then appear in a susceptible material. The criteria adopted as being indicative of a true yield point were that the onset of plastic flow should be abrupt and accompanied by a fall in stress, and that the yield point should be absent in a freshly overstrained specimen and should return on ageing or annealing. Together, these distinguish the yield point phenomenon from other effects which have some similar features, e.g. hardening produced by surface oxidation (Andrade and Randall 1948) and geometrical softening produced when the glide planes are orientated at large angles to the axis of extension (Boas and Schmid 1929, Andrade and Roscoe 1937).

§ 3. ATTEMPTS TO PRODUCE YIELD POINTS

Since it has been shown that crystals of cadmium grown in nitrogen possess yield points (Cottrell and Gibbons 1948), the first attempts to produce yield points in zinc were made by growing crystals in various gases. Nitrogen, carbon dioxide, hydrogen and argon were tried, but in the case of pure zinc no yield points could

be developed. Smooth stress-strain curves, with a critical shear stress of 20–30 gm. mm⁻², were always observed even after repeated strain ageing treatments. Curves 1 and 2 of Figure 1 are examples taken from a crystal grown in nitrogen.

Similar experiments with commercial zinc showed that yield points could always be produced irrespective of the gas in which the crystals were grown. Two strain ageing treatments were usually necessary before the yield point became pronounced. Curves 3, 4 and 5 of Figure 1, taken from a crystal grown in argon, show the formation of a yield point after two strain-ageing treatments. Once formed, the yield point occurred repeatedly after each successive treatment, except when the specimen was disturbed by dismounting it from the machine. The removal of the yield point by plastic overstrain is illustrated by curve 5, which

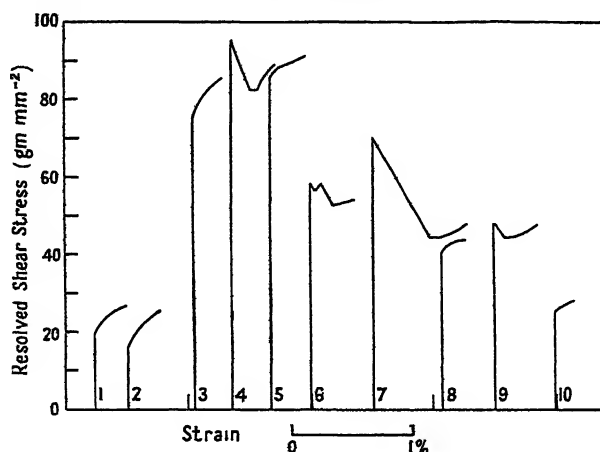


Figure 1. Stress-strain curves on zinc crystals.

- Curve 1. Initial test on pure zinc, crystal grown in nitrogen.
- „ 2. As curve 1, after four strain-ageing treatments.
- „ 3. Initial test on commercial zinc, crystal grown in argon.
- „ 4. As curve 3, after two strain-ageing treatments.
- „ 5. Taken immediately after curve 4.
- „ 6. Commercial zinc, partly degassified, crystal grown in argon, after two strain-ageing treatments.
- „ 7. Pure zinc, melted in air with flux, crystal grown in nitrogen, after three strain-ageing treatments.
- „ 8. Taken immediately after curve 7.
- „ 9. Pure zinc, treated with nitrogen, crystal grown in nitrogen, after one strain-ageing treatment.
- „ 10. Pure zinc, melted in argon with flux, crystal grown in argon, after four strain-ageing treatments.

was taken immediately after curve 4. The upper yield point usually occurred at a resolved shear stress of 80–110 gm.mm⁻² and the lower yield point at 70–95 gm.mm⁻².

These experiments showed not only that yield points could be produced in zinc crystals but also that the effect was associated with impurities in the metal. To identify the impurity element responsible, some crystals were made from pure zinc to which had been added traces of other metals, e.g. 0.01% cadmium. No yield points could be developed, however, and so attention was next directed to gaseous impurities. A small ingot of commercial zinc was melted and solidified several times, *in vacuo*, in an attempt to degassify the metal. Crystals grown in argon from wires extruded from this ingot showed yield points (curve 6,

Figure 1), although these were less marked and the yield stress was reduced to 50–60 gm.mm⁻². This indicated that the effect was related to gas in the metal. It was known that the commercial zinc had been melted in air, under a flux of ammonium and zinc chlorides, and it seemed possible that this procedure might have affected the gas content of the metal. Accordingly, a 50 gm. sample of pure zinc was melted in air in an open combustion boat for 30 minutes, during which time its surface was kept clean by sprinkling it occasionally with pure ammonium chloride. The ingot was then extruded to wire and crystals were grown in nitrogen. These showed the yield phenomenon very strikingly, with a fall in stress of 20–30% at the yield point. Curves 7 and 8 of Figure 1 are examples. In general the upper yield point occurred at 60–80 gm.mm⁻² and the lower yield point at 40–50 gm.mm⁻².

During the preliminary melting operation the ammonium chloride appeared to remove the oxide film completely from the molten metal, which would presumably allow the latter to absorb gases, particularly nitrogen, from the atmosphere more easily. On the other hand, it was possible that ammonium chloride or one of its decomposition products had dissolved in the metal. Two experiments showed that the first possibility was more probable. If dry, oxygen-free nitrogen was bubbled through molten pure zinc, with no flux present, yield points could then be developed in crystals of the metal (curve 9, Figure 1). Secondly, yield points could not be developed in crystals made from pure zinc which had been melted under flux in an atmosphere of argon (curve 10, Figure 1).

Zinc is thus similar to cadmium in that single crystals possess yield points when they contain nitrogen. However, whereas sufficient nitrogen to develop the yield point is absorbed by cadmium during growth of the crystals in a nitrogen atmosphere, in the case of zinc nitrogen has to be introduced in a preliminary melting operation.

§ 4. CHARACTERISTICS OF THE YIELD POINT IN ZINC

The fact that in the above experiments the nitrogen was introduced before the metal was extruded almost disposes of the possibility that the yield point might be a result of a special surface condition, such as, for example, the formation of a nitride on the surface. The indifference of the yield phenomenon to the surface condition was confirmed by experiments on crystals of pure zinc containing nitrogen which were electro-polished before being tested. The polishing was carried out in chromic acid solution (Rodda 1943) and removed a layer about 0.025 mm. deep from the surface. The yield point could be developed very clearly in these specimens (Figure 2) which shows, in agreement with theory, that the effect is due to nitrogen within the metal rather than at its surface.

Figure 2 also illustrates the development of the yield point after successive strain-ageing treatments. It is remarkable how closely the yield point can be reproduced, once it has been clearly developed by the early strain-ageing treatments. In curves 6, 8 and 10 the upper yield occurs consistently at 84–85 gm.mm⁻² and the lower yield at 67–68 gm.mm⁻². This is quite different from the behaviour of iron and steel, where the yield point returns at a higher level of stress after each successive treatment (Muir 1906).

The reason for this became apparent after a series of experiments in which a crystal was aged, after overstraining, at room temperature for various periods of time. Figure 3 shows the results. For short times of ageing the elastic

limit decreases due to recovery and the yield point does not appear. Recovery is rapid at first and after two to four hours the elastic limit reaches a minimum. It then remains constant until the yield point first appears, after eight to nine hours; ageing for progressively longer times causes the yield point to increase. Interpreted theoretically, two phenomena have to be considered, recovery, involving the

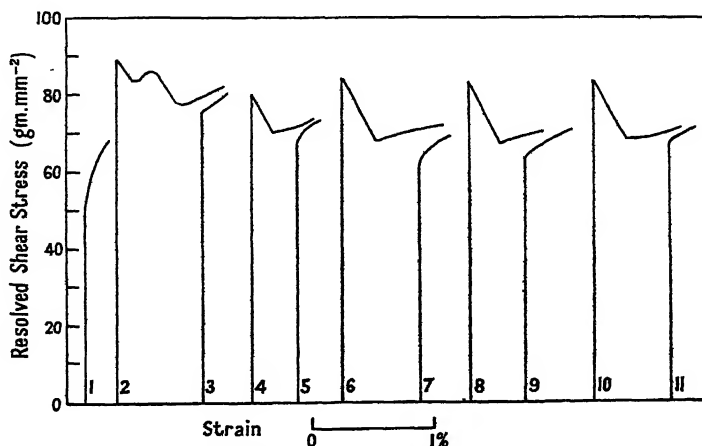


Figure 2. Stress-strain curves on an electrolytically polished crystal made from pure zinc melted in air under ammonium chloride.

- Curve 1. Initial test.
 " 2. After one strain-ageing treatment.
 " 3. Taken immediately after curve 2.
 " 4. After two strain-ageing treatments.
 " 5. Taken immediately after curve 4.
 " 6. After three strain-ageing treatments.
 " 7. Taken immediately after curve 6.
 " 8. After four strain-ageing treatments.
 " 9. Taken immediately after curve 8.
 " 10. After seven strain-ageing treatments.
 " 11. Taken immediately after curve 10.

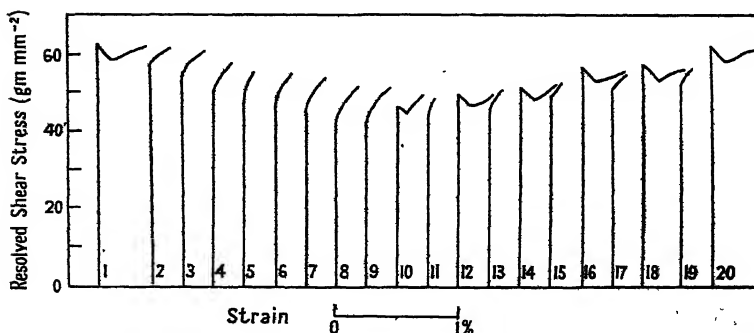


Figure 3. Stress-strain curves, after various ageing treatments at room temperature, on a crystal made from pure zinc melted in air under ammonium chloride.

Curve No.	1	2	3	4	5	6	7	8	9	10
Time of ageing (hours)	48	0	0.25	0.5	1	2	3.75	6	8	8.5
Curve No.	11	12	13	14	15	16	17	18	19	20*
Time of ageing (hours)	0	10	0	11	0	17	0	24	0	48

* Curve 1 repeated.

movement and annihilation of dislocations, and strain ageing, involving the migration of nitrogen atoms to dislocations. It is hardly possible for strain ageing to occur during the early stages of recovery when the dislocations are moving rapidly; however, when recovery is almost completed the solute atoms have a reasonable chance of migrating to the dislocations and fixing them in position. Further evidence on this point was given by ageing a specimen while it was subjected to an applied stress. It was not possible to make the yield point return except when the stress was well below the lower yield value. Higher stresses evidently cause creep by the continued movement of freed dislocations, so that strain ageing can hardly occur.

The rapid recovery of these specimens can be attributed to the high purity of the metal, which allows free dislocations to be highly mobile. In the case of iron, however, the metal is much less pure and it is well known that thermal softening only becomes rapid at temperatures approaching 500°. As a result, strain ageing occurs before the strain hardening is removed; the increase in the yield point, with successive straining and ageing treatments, can thus be traced to the accumulation of strain hardening in the material. In pure zinc, on the other hand, recovery occurs almost completely during each strain-ageing treatment so that the yield point returns at the same level.

The experiments summarized in Figure 3 were repeated, using slightly elevated ageing temperatures, and showed that the rate of ageing increased with temperature. Thus, at 30, 40 and 50°C., the times taken for the first appearance of the yield point were 140, 45 and 15 minutes, respectively. It may be deduced from these results that the process is thermally activated, with an activation energy of 21,000 cal. per mol.

Smith (1947) reported that the magnitude of the yield point in strain-aged cadmium crystals depended upon λ , the angle between the axis of tension and the glide direction, being very small when this angle was about 40–45°; this was not confirmed by Cottrell and Gibbons (1948) who obtained large yield points in cadmium crystals of all orientations. In the present work no dependence of the yield point of zinc on orientation was found although it was noticed that more strain-ageing treatments were usually needed to develop the yield point fully in crystals with $\lambda \simeq 45^\circ$ than in others. For a given critical shear stress, a crystal having this orientation is softer and more easily overstrained than any other; it is possible that more strain-ageing treatments were needed because such specimens are more easily bent during preparation and mounting prior to testing.

§ 5. DISCUSSION

In each of the four metals iron, molybdenum, cadmium and zinc in which the yield phenomenon has been clearly established, it has been shown that nitrogen is an effective solute element for producing the yield point. Carbon is also effective in iron, but its behaviour in the other metals has not so far been studied. Theoretically, the conditions for a solute atom to produce the yield phenomenon are that it should distort the lattice severely and be able to migrate fairly rapidly, conditions which are known to be fulfilled in the case of carbon and nitrogen in iron. The behaviour of these elements in the other metals has not been sufficiently studied, but it seems probable that they dissolve interstitially, in which case they may be expected both to distort the lattice and to be highly mobile.

In no metal of face-centred cubic structure has a yield point been established so clearly as in the above metals. Frank (private communication) has suggested that the comparative rarity of the effect may be because most solute atoms probably interact weakly with screw dislocations. The hydrostatic stress round a screw dislocation is ideally zero, so that the dislocation may not attract solute atoms that produce only spherically symmetrical distortions. Solute atoms that produce non-symmetrical distortions are needed to anchor this type of dislocation, since these can interact with its shear stress field. Carbon and nitrogen are known to produce this type of distortion in α -iron and probably in other body-centred cubic transition metals (Dijkstra 1947, Kê 1948), but not in face-centred cubic metals; this is a consequence of the shapes of the interstitial positions occupied by these atoms in the two types of lattice. In the case of the close packed hexagonal structure the presence of a crystallographically unique axis obviously favours a non-symmetrical distortion round solute atoms.

Strongly marked yield points may therefore be rare in face-centred cubic metals because of the difficulty of anchoring screw dislocations in this type of lattice. This does not mean, of course, that edge dislocations will not be anchored; the strong hydrostatic component in their stress fields should enable them to attract many types of solute atoms, but this can hardly lead to a yield point while the screw dislocations remain free.

ACKNOWLEDGMENTS

The authors wish to thank Professor D. Hanson, under whose general supervision this work was carried out, for his interest and support, and Mr. D. F. Gibbons and Dr. A. T. Churchman for many useful discussions. The spectrographic analyses were kindly made by the laboratories of the Imperial Smelting Corporation.

REFERENCES

- ANDRADE, E. N. DA C., and RANDALL, R. F. Y., 1948, *Nature, Lond.*, **162**, 890.
 ANDRADE, E. N. DA C., and ROSCOE, R., 1937, *Proc. Phys. Soc.*, **49**, 152.
 BARRETT, C. S., and LEVENSON, L. H., 1940, *Trans. Amer. Inst. Min. Met. Engrs.*, **137**, 76, 112.
 BOAS, W., and SCHMID, E., 1929, *Z. Phys.*, **54**, 16.
 COTTRELL, A. H., 1948, *Report on the Strength of Solids* (London: Physical Society), p. 30.
 COTTRELL, A. H., and BILBY, B. A., 1949, *Proc. Phys. Soc. A*, **62**, 49.
 COTTRELL, A. H., and GIBBONS, D. F., 1948, *Nature, Lond.*, **162**, 488.
 DIJKSTRA, L. J., 1947, *Philips' Res. Rep.*, **2**, 357.
 EDWARDS, C. A., PHILLIPS, D. L., and JONES, H. N., 1940, *J. Iron and Steel Inst.*, **142**, 199.
 Kê, T. S., 1948, *Phys. Rev.*, **74**, 9.
 LOW, J. R., and GENSAMER, M., 1944, *Trans. Amer. Inst. Min. Met. Engrs.*, **158**, 207.
 MUIR, J., 1906, *Proc. Roy. Soc. A*, **77**, 277.
 NABARRO, F. R. N., 1948, *Report on the Strength of Solids* (London: Physical Society), p. 38.
 OROWAN, E., 1934, *Z. Phys.*, **89**, 634; 1940, *Proc. Phys. Soc.*, **52**, 14.
 POLANYI, M., 1925, *Z. Tech. Phys.*, **6**, 121.
 RODDA, J. L., 1943, *Min. and Met.*, **43**, 323.
 SMITH, C. L., 1947, *Nature, Lond.*, **160**, 466.
 SNOEK, J. L., 1941, *Physica*, **8**, 734.
 TÚRY, P., and KRAUSZ, S., 1936, *Nature, Lond.*, **138**, 331; 1937, *Ibid.*, **139**, 30.

A Simple Constant Stress Apparatus for Creep Testing

By L. M. T. HOPKIN*

British Non-Ferrous Metals Research Association, London

Communicated by A. G. Quarrell; MS. received 16th November 1949

ABSTRACT. A simple device is described which can maintain the stress on a creep specimen constant to within 0.8% during uniform extensions up to 100%. The apparatus is suitable for slow rates of strain. Examples are given of creep curves obtained with this device from tests on lead and a lead-tin alloy extending to 1,000 and 400 hours respectively; good agreement with the Andrade creep equation is observed in both cases.

LONG time creep tests are usually performed under conditions of constant load for the sake of simplicity, although this method has the disadvantage that the stress increases as the cross-sectional area of the specimen decreases during extension. The change in stress is insignificant when the extension is small, but it becomes important in tests on materials which undergo large extensions before failure. In consequence Andrade, in his classic work on the creep of metals (Andrade 1910), carried out his tests at constant stress and was able to derive a general equation for the creep curves so obtained. Most subsequent fundamental investigations of the phenomenon of creep in ductile materials have been carried out under conditions of constant stress.

In a long-term investigation of the influence of metallurgical factors on the creep of lead and lead alloys now in progress in the laboratories of the British Non-Ferrous Metals Research Association, the tests are being made at constant stress to simplify interpretation of the results and to enable the creep curves to be analysed mathematically. The results obtained should increase the knowledge of the mechanism of flow in metals as it will be possible, for example, to test the validity of the Andrade equation in creep tests of long duration. In this investigation up to 150 specimens will be tested simultaneously and many of the tests will be continued for several years. It was therefore essential that the constant stress device employed should be simple in construction, relatively inexpensive, and suitable for slow rates of strain by being as frictionless as possible in operation. These considerations ruled out the designs which have already been described in the literature (Andrade 1910, 1914, 1948, Andrade and Chalmers 1932, Pearson 1934, Fisher and Carrker 1949, Ward and Marriott 1948), but a simple apparatus has been designed and tested which fulfils all requirements.

The apparatus is shown in Figure 1, and is a modification of the lever used by Andrade and Chalmers (1932). The method of suspension by knife edges used by these workers has been replaced by an arrangement of steel tapes. Figure 2 shows a creep testing frame containing specimens of lead and lead alloys loaded by means of these constant stress levers.

As the specimen extends, the moment of the applied weights about a point in the line of action of the anchoring tapes is inversely proportional to the length of the specimen, and hence directly proportional to the cross-sectional area of the specimen assuming the specimen does not change its volume during extension. Like other constant stress devices, this new apparatus maintains a constant stress only if the specimen extends uniformly.

* Investigator, B.N.F.M.R.A., London.

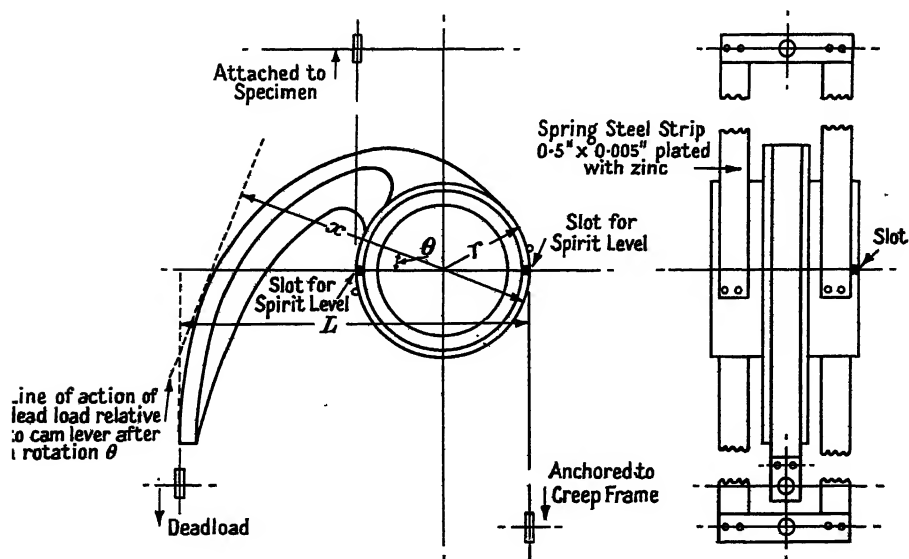


Figure 1. Constant stress cam lever.

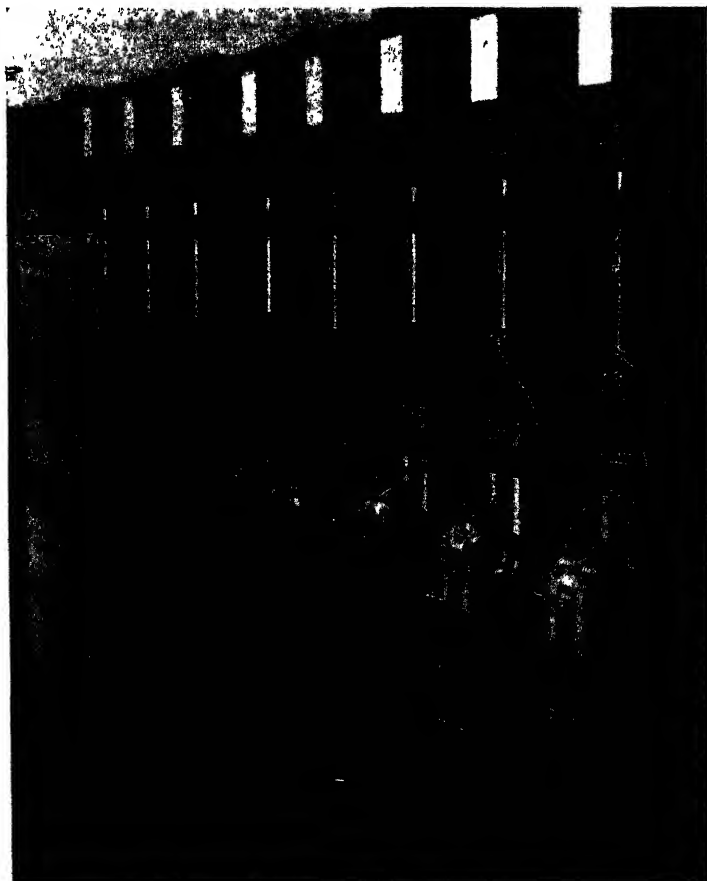


Figure 2. Creep frame containing specimens of lead and lead alloys loaded by means of the constant stress levers.

The shape of the cam of the lever was determined graphically from the relationship

$$\frac{x}{L} = \frac{a}{a+2r\theta}, \quad \dots\dots(1)$$

where a is the initial length of the specimen, θ is the angular rotation of the lever, r is the radius of the cylindrical drum of the lever, and x is the arm of the moment of the weights about a point in the line of action of the anchoring tapes after an angular rotation of θ ; L is the initial arm of the moment when $\theta=0^\circ$.

Equation (1) is similar to that given by Andrade and Chalmers except for the factor 2 in the denominator. This is introduced because, in the present case, the effective fulcrum falls as the specimen extends.

The size of the lever was determined from the dimensions $L=6$ in., $a=7$ in. and $r=1.5$ in., which permit the stress on a specimen to be maintained constant for uniform extensions up to 100%. After 100% uniform extension the test continues under constant load. It will be noted that the initial lever ratio is 2:1 when $\theta=0$.

The lever, which was designed so that the maximum initial force of 225 lb. wt. could be exerted on a specimen, consists of a single casting of a high strength aluminium alloy machined to the required dimensions. The connections from the lever to the weights, to the specimen, and to the point of anchor on the creep testing frame are made with spring steel tape, 0.5 in. wide and 0.005 in. thick, electrolytically plated with zinc, 0.0005 in. thick on each side, to prevent corrosion. Any hydrogen embrittlement of the tapes which may have resulted from the plating process was removed by a suitable heat-treatment. The weight of a complete lever is 1 lb.

For accurate and reproducible results the inclination of the centre line of the lever to the horizontal at the beginning of a test must be known and capable of adjustment. This inclination is determined by a small spirit level mounted so that it can be readily fitted, when required, into two slots milled on the horizontal centre line of the apparatus (see Figure 1). An adjustable anchorage for the tapes on the creep-testing frame enables the inclination of the lever to be adjusted as required.

When arranging the inclination of the lever relative to the horizontal at the beginning of a test, allowance is made for the rotation of the lever which occurs on loading as the result of the elastic extension of the tapes. The relationship between the rotation and the applied load can be determined experimentally as follows, and is the same for all the levers. The lever is set up as it is used in practice with the exception that the creep specimen is replaced by a steel bar of such dimensions that its elastic extension is negligible under the maximum load used in the creep tests. A suitable load is attached to the lever and the anchoring connection is adjusted until the centre line of the lever is horizontal. The load is removed with the exception of a small weight sufficient to keep the tapes taut, and the new position of the bubble of the spirit level is marked. This procedure is repeated for several of the loads to be used in the creep tests. When starting a creep test the lever, loaded with the same small weight sufficient to keep the tapes taut, is adjusted so that the bubble of the spirit level is at the mark corresponding to the load to be used for the particular creep test.

The major error of the device arises from the force applied to the specimen by the moment of the weight of the casting about the effective fulcrum. This error, together with any others, can be determined by calibrating the lever in a tensile:

testing machine. Over the range of forces of 100–225 lb. wt. which could be applied to a specimen at the beginning of a test, it was found that at any strain up to 100% the error in the force would not be greater than 0.8% of the theoretical value. For initial forces between 40–100 lb. wt. it was found that at any strain up to 15% the error in the force on a specimen would not be greater than 0.3% of the theoretical value, although for strains between 15% and 100% the error increased to 0.8%.

If necessary, the error due to the weight of the lever can be minimized for small strains by making suitable additions to weights to the body of the lever so that the centre of gravity of the apparatus falls near the effective fulcrum at the beginning of a test.

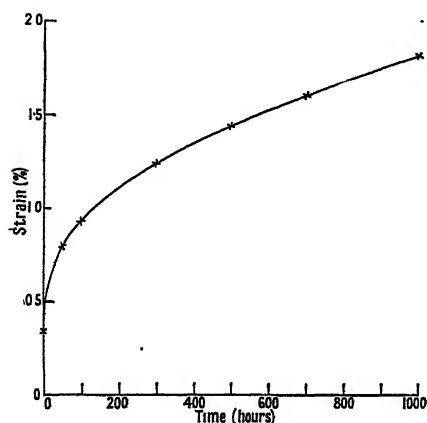


Figure 3. A creep test at a constant stress of 300 lb/in² on lead of 99.999% purity.

Experimental curve shown by continuous line. The points shown by crosses were calculated from the constants:

$$l_0 = 1.0034, \quad \beta = 1.2 \times 10^{-3} \text{ hr}^{-1/3}, \\ K = 3.0 \times 10^{-6} \text{ hr}^{-1}.$$

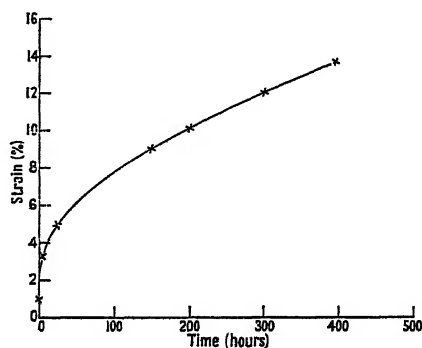


Figure 4. A creep test at a constant stress of 1,000 lb/in² on an alloy of lead containing 1% of tin.

Experimental curve shown by continuous line. The points shown by the crosses were calculated from the constants:

$$l_0 = 1.0097, \quad \beta = 1.29 \times 10^{-3} \text{ hr}^{-1/3}, \\ K = 7.4 \times 10^{-6} \text{ hr}^{-1}.$$

The creep curves shown in Figures 3 and 4 were obtained from tests carried out on lead of 99.999% purity and the same lead alloyed with 1% of tin using the constant stress device. In both figures the continuous line passes through the experimentally determined points, which have been omitted for the sake of clarity. The crosses indicate the points calculated from the Andrade equation $l_t = l_0(1 + \beta t^{1/3})e^{Kt}$ where l_t is the length of the specimen at time t and l_0 , β (hr^{-1/3}) and K (hr⁻¹) are constants.

ACKNOWLEDGMENTS

The author wishes to thank the Director and the Council of the British Non Ferrous Metals Research Association for permission to publish this paper. The method of suspension of the lever was evolved from a suggestion made by Mr. R. May, A.R.S.M.

REFERENCES

- ANDRADE, E. N. DA C., 1910, *Proc. Roy. Soc. A*, **84**, 1; 1914, *Ibid.*, **90**, 329; 1948, *Proc. Phys. Soc.*, **60**, 304.
- ANDRADE, E. N. DA C., and CHALMERS, B., 1932, *Proc. Roy. Soc. A*, **138**, 348.
- FISHER, J. C., and CARRKER, R. P., 1949, *J. Metals*, **1**, 178.
- PEARSON, C. E., 1934, *J. Inst. Metals*, **54**, 111.
- WARD, A. G., and MARRIOTT, R. R., 1948, *J. Sci. Instrum.*, **25**, 147.

Surface Tensions in the System Solid Copper-Molten Lead

BY G. L. J. BAILEY AND H. C. WATKINS*

The British Non-Ferrous Metals Research Association, London

MS. received 21st October 1949

ABSTRACT. A method is described by which the surface tension of surfaces occurring in the systems copper-hydrogen or argon and copper-lead-hydrogen or argon is determined. At temperatures round 850° C. the surface tension between copper grains is 640 dyne/cm. and between copper and the equilibrium lead-copper liquid phase 340 dyne/cm. The surface tension of a copper-gas surface is the same whether the gas is hydrogen or argon, within the limits of precision of the experiment, and has the value 1,800 dyne/cm. When lead vapour is present in the gas atmosphere at a partial pressure of the order of 0.1 mm. Hg the surface tension of the copper-gas surface has the much lower value of 780 dyne/cm. It is suggested that the reduction is due to the adsorption of lead at the copper-gas surface. The failure of the liquid phase to exhibit a contact angle of zero against solid copper surfaces is thereby understood.

§ 1. INTRODUCTION

THE wetting of a solid by a liquid may be regarded formally as being determined by the surface tensions at the three surfaces solid-liquid, liquid-gas and solid-gas. If the contact angle exceeds zero its value shows the balance of these three tensions, but generally the absolute magnitude of only one of them, the liquid-gas surface tension, is known or readily ascertainable. The present paper describes an attempt to measure the two unknown surface tensions using methods inspired by the work of C. S. Smith (1948) on the rôle of surface tensions in determining the structure of fully annealed alloys.

In the course of the British Non-Ferrous Metals Research Association's investigation of soldering and brazing, simple experiments were made in which the angle at which the surface of a liquid lead-copper alloy met a solid surface of copper was estimated, at a series of temperatures up to 950° C., the metals being heated in an atmosphere of purified cylinder hydrogen. It was noted that, especially at the higher temperatures employed, the grain-boundaries of the exposed copper surfaces rapidly became etched and filled with liquid. Further experiments, in which electropolished specimens of oxygen-free copper were heated in an atmosphere of hydrogen or argon with or without the presence of unsaturated lead vapour showed that marked thermal etching of the grain-boundaries occurred when lead vapour was present, the effect in hydrogen or argon alone being comparatively small.

The interpretation placed upon these observations was that proposed by Chalmers, King and Shuttleworth (1948), that the development of grain-boundary grooves represents an approach to an equilibrium condition of minimum surface free energy, the mechanism probably being that of surface migration of copper ions.

Surface tension is often regarded as equivalent, and numerically equal, to specific surface free energy. Gurney (1949) has recently discussed the conditions under which such an assumption is justified. In his view surface

* Now with Messrs. H. J. Enthoven & Sons, Ltd., London.

tension and specific surface free energy are numerically equal only when surface migration of atoms has occurred to the extent required for the establishment of thermal equilibrium, so that both are minimal. In the experiments to be described, the attainment of equilibrium was always sought. The terms surface tension and specific surface free energy are therefore taken to be equivalent.

It seemed likely that the surface tension of copper is lower in the presence of lead vapour than in the presence of argon or hydrogen alone, since the boundary grooves were deeper. Such a decrease might possibly result from the adsorption of lead from the vapour phase, but, whatever the reason, it could explain why the contact angle of lead-copper liquid against a copper surface is greater than zero. The value of the contact angle is determined by the surface tensions of the solid-gas, liquid-gas and solid-liquid interfaces. The surface tension of a clean copper surface in contact with its own vapour and hydrogen is expected to be greater than 1,000 dyne/cm., since that of liquid copper just above the melting point has been found to exceed this value. The liquid-gas surface tension is not likely to be greater than 500 dyne/cm., and, according to Smith (1948), the solid-liquid surface tension is about one-half that of the surfaces between grains of the solid, which is itself several times smaller than that of the solid-gas surface. Thus the liquid phase would be expected always to exhibit a contact angle of zero against a copper surface cleaned in hydrogen. This is not, in fact, observed.

It appeared that absolute values for the surface tension of copper in hydrogen or argon, and also of copper in hydrogen or argon containing lead vapour, might be determined, together with values for the copper-lead surface and for the surface between grains in solid copper. These would be derived, in terms of the surface tension of the liquid phase, from measurements of the angles at the root of the etched grain-boundaries and of the contact angle between the liquid lead-copper alloy and copper, together with some of the measurements quoted by Smith. The derivation of these absolute values is described below.

§ 2. PRINCIPLE OF THE METHOD

Smith (1948) has proposed that the grain structure of many well-annealed alloys reveals an approach to equilibrium between phase and grain interfaces whose surface tensions balance each other at the points and along the lines where they meet. In a particular case of a well-annealed two-phase alloy, for example of copper containing a small percentage of lead, a normal cross section of a boundary between three copper grains might appear as in Figure 1, the lead being present as a prismatic filling along the edges where the copper grains would otherwise meet. At each corner of the prism the balance of surface tensions is represented by the equation

$$\gamma_{11} = 2\gamma_{12} \cos \theta / 2 \quad \dots \dots (1)$$

in which γ_{11} is the surface tension between copper grains, γ_{12} the surface tension between copper and lead phases, and the angle θ is called by Smith the 'dihedral angle'. Smith suggests that the values of γ_{11} and γ_{12} are substantially independent of crystallographic orientation and, from measurements of the dihedral angle θ , quotes values for the ratio γ_{12}/γ_{11} of a number of alloys at several temperatures.

The equilibrium configuration of grain-boundaries exposed in a solid-gas surface is similarly determined by the balance of surface tensions (Chalmers, King and Shuttleworth 1948). Figure 2 represents a normal cross section

through a grain-boundary groove produced by thermal etching, for example of copper. The balance of surface tensions is given by the equation

$$\gamma_{11} = 2\gamma_{13} \cos \phi/2 \quad \dots\dots (2)$$

in which γ_{11} is the surface tension between copper grains, γ_{13} the surface tension between copper and gas phase, and ϕ the dihedral angle.

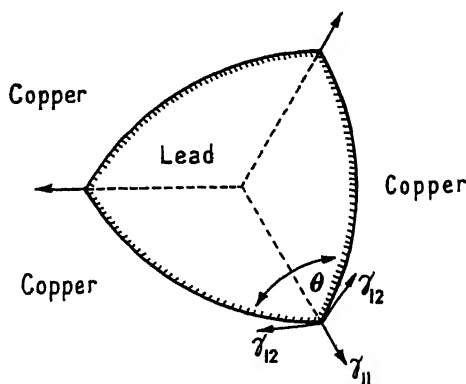


Figure 1.

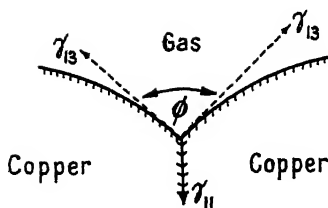


Figure 2.

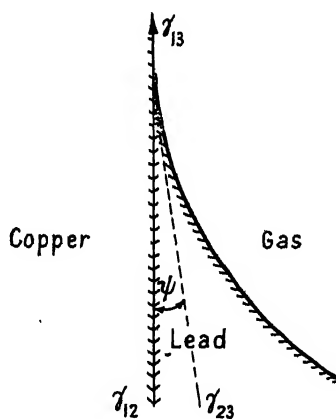


Figure 3.

It is assumed that γ_{13} is independent of relative orientation.

From Figure 3 it is clear that the balance of surface tension forces determining the contact angle exhibited against a smooth solid surface is given by the expression

$$\gamma_{13} = \gamma_{12} + \gamma_{23} \cos \psi \quad \dots\dots (3)$$

where γ_{23} is the surface tension of lead-copper liquid, ψ the angle of contact, and γ_{12} , γ_{13} have the meanings previously assigned.

The values of γ_{11} , γ_{12} , γ_{13} , γ_{23} must all be considered to depend on temperature, and, since adsorption may have important effects, on the components of the gas phase and their partial pressures. It will be assumed that γ_{11} (copper-copper), γ_{12} (copper-lead) and γ_{23} (lead-gas) are independent of the nature of the gas phase, which affects only the solid-gas surface tension, γ_{13} . Then, from measurements of the angles θ , ϕ and ψ , the values of γ_{11} , γ_{12} and γ_{13} may be readily determined in terms of the surface tension of the liquid lead-copper alloy, γ_{23} .

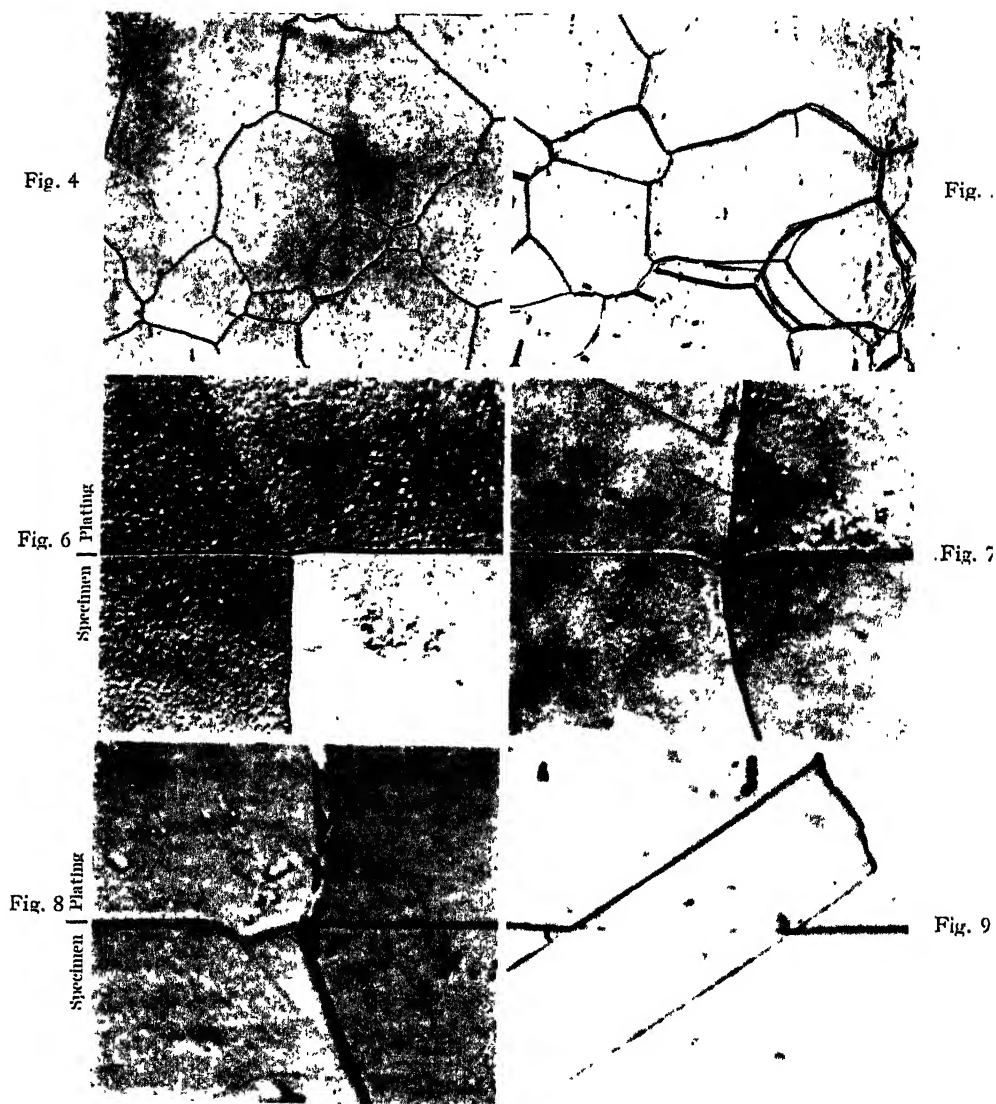


Figure 4. Grain boundary etching of electrolytically polished copper surface heated in argon for two hours at 900° C. $\times 50$.

Figure 5. Grain boundary etching of electrolytically polished copper surface heated in argon with unsaturated lead vapour for two hours at 900° C. $\times 50$.

Figure 6. Section of groove produced at grain-boundary by thermally etching electrolytically polished copper in hydrogen for two hours at 900° C. $\times 2500$.

Figure 7. Section of groove produced at grain-boundary by thermally etching electrolytically polished copper in hydrogen with unsaturated lead vapour for two hours at 900° C. $\times 2500$.

Figure 8. Same treatment as for Figure 7. Section showing detached groove and grain-boundary. $\times 2500$.

Figure 9. Same treatment as for Figure 7. Section showing intersection of twin boundaries with surface. $\times 2500$.

§ 3. EXPERIMENTAL

(i) *Thermal Etching Technique*

The materials used were oxygen-free high-conductivity copper and high-purity (99.998%) lead. The copper specimens, 25 mm. \times 5 mm., cut from 16-gauge sheet, were electrolytically polished in phosphoric acid after preliminary polishing on emery papers. After this treatment the copper surface appeared smooth and featureless when examined microscopically at $\times 2,500$.

Annealing was carried out in a silica tube heated in a furnace which could be withdrawn after the experiment so as to promote rapid cooling from the temperature of treatment. The furnace atmosphere was either cylinder hydrogen or argon at atmospheric pressure, freed from residual oxygen by passage over heated platinized asbestos, followed by a series of drying tubes and further through a second similar train. When argon was used the specimens were subjected to a de-oxidizing treatment in hydrogen before switching over to the argon supply.

Preliminary experiments indicated that the development of grain-boundary grooves in hydrogen atmosphere was rather slow at temperatures below 800°C., and the investigation was confined to a relatively narrow temperature range, viz. 800–900°C.

In the experiments using an atmosphere containing lead vapour the lead was contained in a small silica boat placed in the silica tube upstream of the copper specimen. It was arranged that when the furnace was in position the specimen temperature was always about 50°C. higher than the temperature of the lead, so that the lead vapour was unsaturated above the specimen. No attempt was made in these experiments to control the vapour pressure rigidly. Its value was taken to be of the order of 0.1 mm. Hg.

After the grain-boundary etching treatment the specimens were cooled by withdrawing the furnace, and the surfaces were protected by plating in an acid copper bath. The plated specimens were sectioned, mounted in bakelite, and polished in the normal manner for metallographic examination. The best etch was found to be a mixture of equal volumes of 0.88 ammonium hydroxide and 10 vol. hydrogen peroxide. The surface appearance of boundary grooves developed in the absence and presence of lead vapour is shown in Figures 4 and 5 (see Plate) respectively, and etched sections through typical grooves appear in Figures 6 and 7.

It is of interest to note that the surfaces of specimens exposed to lead vapour afterwards showed little tendency to tarnish in the laboratory atmosphere, in comparison with similar specimens treated in hydrogen only.

(ii) *Measurement of Dihedral Angle ϕ (Solid-Gas)*

Dihedral angles of the proper value will be observed on a section through the surface only if the following conditions are met:

- (a) The plane of the microsection is normal to the surface of the specimen.
- (b) The plane of the grain boundary is perpendicular to the surface of the specimen.
- (c) The plane of the grain boundary is perpendicular to the plane of the microsection.

Care was taken to cut sections perpendicular to the surface of the specimen, and no serious error is likely to have been introduced from this cause. It was observed that most of the boundaries were substantially perpendicular to the surface of the specimen. Any showing deviations of more than a few degrees were neglected in the measurement of dihedral angle.

Compliance with condition (c) above is clearly impracticable when polycrystalline specimens are used. The effect of a departure from perpendicularity between the plane of the grain-boundary and that of the microsection is to enlarge the dihedral angle; thus to obtain a true value only the smallest observed value should be taken. The directions of the grain-boundaries upon the thermally etched surface are random, so that the probability that the plane of a boundary will lie within a given angular range about a particular angle to the plane of the normal section examined is the same for all angles. By inspection it was found that the lowest angle obtained on any specimen was about 120° . With a true dihedral angle of this value it may readily be shown that, if the plane of the boundary is within $\pm 22.5^\circ$ of the normal to the plane of section, the apparent dihedral angle does not exceed the true value by more than 3° , and the probability that a given angle is within 3° of the true value is thus 1:4. The error of measurement was expected to be of this order.

It is reasonable to suppose that the equilibrium dihedral angle, ϕ , is established as soon as thermal etching begins, and that, as etching proceeds, the grooves merely become deeper, while maintaining the same root angle. The flanks of a groove are cylindrically curved surfaces, intersecting at its root, and the measurement of root angle on a section requires that tangents shall be drawn to the traces of these surfaces at their point of intersection. The deeper the groove, the more accurate is this process likely to be. However, the depth of grooves did not increase proportionately to time of anneal after the first two hours. The reason for this is probably that grain-growth took place during annealing, so that boundaries were in general moving slowly and equilibrium tended to be established between the rate of boundary migration and depth of groove, determined by the rate of surface migration of copper atoms. A sudden increase in rate of boundary migration could supposedly result in a boundary becoming detached from its groove, which should then begin to heal. Such a detached groove and boundary are shown in section in Figure 8, and surface scars resulting from a change in rate of movement may be seen in Figure 5. The difficulty could no doubt be overcome, for example, by using bicrystal specimens; by this means also the dependence of surface tensions on crystallographic orientation could be ascertained.

However long the annealing time, the depth of grooves did not exceed that typified in Figures 6 and 7, in which the magnification is 2,500 times.

Tangents at the root of selected grooves were drawn in by eye on prints of such micrographs enlarged a further five times.

A best value for the dihedral angle was derived in the following way. A large proportion of the boundaries observed were between twins, and did not develop grooves (Figure 9), and some of the remaining boundaries were not perpendicular to the surface of the specimen. These were all discarded. Of the remainder, some were not deep enough to be measurable reliably, or were ill-defined, but on most of them measurement could have been attempted. From all these a selection was made of that quarter of the total appearing to have the

lowest angles. These were photographed and measured as described above. It was assumed that in all these cases the plane of the boundary was normal to that of the section, and the most frequently occurring angle was taken as the best value for the true angle.

An analysis of observations made on a typical specimen is given in Table 1, and in Table 2 are set out the values of the angle measured for each of the twelve grooves photographed, as stated in Table 1.

Table 1. Analysis of Observations on Typical Specimen
(Copper heated for 8 hours at 800° c. in hydrogen with unsaturated lead vapour)

Discarded boundaries :	
Twin boundaries	154
Boundaries not perpendicular to surface of specimen	12
Selected boundaries :	
Dihedral angles not measurable or too large	34
Dihedral angles photographed	12

Table 2. Measured Values of Dihedral Angle on Typical Specimen

Number	1	2	3	4	5	6	7	8	9	10	11	12
Angle (deg.)	119	123	124	128	128	128	129	129	132	132	144	157

From the greater frequency of occurrence of the angles 128–129°, it was considered that the value 129° could be taken as the true dihedral angle in this case. A similar procedure was applied in the remaining cases, and the results are summarized in Table 3.

Table 3. Estimated Values of Dihedral Angle, ϕ

Gas phase	Temperature (° c.)	Annealing time (hr.)	Dihedral angle (deg.)
Argon	900	2	160
Hydrogen	800	14	158
	900	2	160
Lead vapour-Argon	900	2	142
Lead vapour-Hydrogen	800	8	129
	900	2	129

To enable these values of dihedral angle ϕ to be used in calculating absolute values for the surface tensions, estimates are required of the value, at 800° and 900° c., of the microstructural dihedral angle θ , as measured by Smith, and also of the contact angle ψ , and of the surface tension of liquid lead containing equilibrium proportions of copper. These estimates are made below.

(iii) *Estimation of Dihedral Angle θ , (Solid-Liquid)*

The following values of the dihedral angle θ are given by Smith for two-phase copper-lead alloys for temperatures 600°, 700° and 900° c. respectively: 70°, 65°, 50°; by interpolation the angle at 800° c. is 57°.

(iv) *Estimation of Surface Tension of Liquid Phase, γ_{23}*

Values of the surface tension of lead over a range of temperatures have been given by Bircumshaw (1934) and by Sauerwald and Drath (1927). Values for copper between 1,131° and 1,215°C. are also given by Sauerwald and Drath. From an extrapolation of the values for copper to the temperatures of interest the surface tensions of the liquid phase at 800° and 900°C. are estimated, using a simple mixtures rule, to be 435 and 466 dyne/cm. respectively.

(v) *Estimation of Contact Angle ψ (Solid-Liquid-Gas)*

The contact angle ψ is that angle which would be observed at the line of contact of liquid and solid surfaces if the solid surface were perfectly smooth. Real surfaces are seldom smooth, however, and, according to Wenzel (1936), when the ratio of actual surface area to the nominal area is r the angle actually established is λ , where $\cos \lambda = r \cos \psi$.

It has been shown by Shuttleworth and Bailey (1948), however, that this is true only when the roughness takes the form of interlacing grooves. When it has the character of isolated depressions, the observed contact angle is no longer uniquely determined by the roughness. When the liquid advances over the surface the contact angle exceeds the true value, and when it recedes the observed angle is less than the true value, i.e. hysteresis of the contact angle is exhibited. Hysteresis is normally encountered in metallic systems, in which intermetallic diffusion frequently results in the roughening of initially polished surfaces.

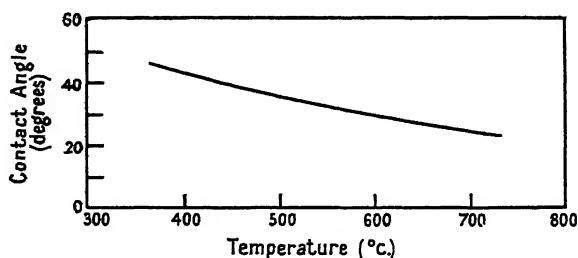


Figure 10. Contact angle vs. temperature: copper-lead system.

Even when electropolished copper is dipped into a liquid copper-lead alloy of equilibrium composition at the temperature of experiment some roughening rapidly occurs at the solid-liquid interface and hysteresis is observed, although the unwetted surface remains smooth for a longer time. It is probable, however, that the advancing contact angle is in this case near to the true value. Measurements have been made of this angle, using the well-known Wilhelmy plate method (Wilhelmy 1863), in the temperature range 350–750°C., the furnace atmosphere being purified hydrogen. The results are shown in Figure 10. Observation by eye at temperatures up to 950°C. suggests that it is reasonable to extrapolate the curve of Figure 10 to estimate values for the contact angle at 800° and 900°C.; these estimated values are 25° and 15° respectively. Observation by eye of the angle exhibited in an argon atmosphere shows that it cannot differ greatly from the values at corresponding temperatures when a hydrogen atmosphere is used and these estimates are therefore used for both conditions.

The use of the advancing contact angle is appropriate in view of the possible occurrence of adsorption hysteresis in addition to the roughness hysteresis already mentioned.

§ 4. CALCULATION OF SURFACE TENSIONS

The estimates of angle ϕ , θ and ψ and of the surface tension of the liquid phase at temperatures of 800° and 900° c. are inserted in the appropriate equations, (1) to (3), to calculate values for the surface tensions γ_{11} , γ_{12} and γ_{13} . These are given in Table 4.

There are two numerical values for the solid-gas surface tension of copper, one appropriate to a lead-free atmosphere and one to atmosphere containing lead vapour.

Table 4. Evaluation of Surface Tensions

Temp. (° c.)	Gas phase	γ_{23} (dyne/cm.)	θ (deg.)	ψ (deg.)	ϕ (deg.)	γ_{13} (dyne/cm.)	γ_{12} (dyne/cm.)	γ_{11} (dyne/cm)
800	Hydrogen	—	—	—	158	1750	—	670
	Hydrogen + lead	435	57	25	129	780	380	
900	Hydrogen	—	—	—	160	2000	—	700
	Hydrogen + lead	466	50	15	129	810	360	
900	Argon	—	—	—	160	1600	—	560
	Argon + lead	466	50	15	142	740	290	

§ 5. DISCUSSION OF RESULTS

In view of the assumptions and approximations made, the accuracy of the values obtained cannot be great, and discussion of the variations of individual surface tensions would be unprofitable. Average values of all the results are as follows:

Clean copper:	1,800 dyne/cm.
Between copper grains:	640 dyne/cm.
Between copper and lead:	340 dyne/cm.
Copper with adsorbed lead:	780 dyne/cm.

The value of 1,800 dyne/cm. obtained for the surface tension of clean copper is in fair agreement with that recently found by Udin, Schaler and Wulff (1949) using another method. Their value is 1,370 dyne/cm. at the melting point of copper. Using their quoted temperature coefficient of -0.46 dyne/cm. per degree, the value at 850° c. would be about 1,500 dyne/cm. A recent theoretical estimate by Huang and Wyllie (1949) gives a value of 1,820 dyne/cm. for solid copper at its melting point. The surface tension between grains of copper is, as expected, much lower than the surface tension of clean copper, though still surprisingly high, and its ratio with that of the copper-lead phase boundary is, of course, that given by Smith. The substantial reduction of the surface tension of a copper surface exposed to lead vapour is evident.

Clearly, lead vapour at a very low partial pressure is extremely effective in this respect, whereas hydrogen, which is soluble in copper, and therefore liable

to become adsorbed, has little effect. The effect of hydrogen alone is little more marked than that of argon alone, where adsorption is not expected and the copper surface should be clean.

From the experimental values of contact angle given in Figure 10 it is possible, with the aid of equation (3) and assuming γ_{12} and γ_{23} to remain constant with temperature, to estimate the surface tension of a copper surface exposed to lead vapour in a hydrogen atmosphere throughout the temperature range 350–750° c. The results are shown in Table 5.

Table 5. Estimated Surface Tension of Copper in presence of Lead Vapour

Temperature (° c.)	350	450	600	750
Surface tension (dyne/cm.)	420	570	640	730

It is interesting to note that at a temperature just above the melting point of lead the surface tension of copper exposed to lead vapour at low pressure is on this basis roughly the same as the surface tension of liquid lead. It seems probable that the lead exerts its effect on the contact angle through adsorption on to the solid surface, and possible that, at temperatures round 350° c., the copper surface is just covered by a complete layer of lead.

§ 6. CONCLUSION

Several approximations and assumptions have been made in deriving the absolute surface tension values quoted. Many of these could doubtless be eliminated by further experiment. The results obtained are, however, of value in pointing to the reason why liquid lead does not spread completely over copper surfaces, and in showing the approximate magnitudes of the surface tensions of the various surfaces.

The failure of lead to spread readily upon copper appears to be a fundamental feature of the lead-copper system, and is not, for example, due to adventitious contamination of the surface.

The results are of wider interest in demonstrating the strongly marked reduction in the surface tension of a solid metal surface which may be brought about by the adsorption upon it of atoms of another metal, and it follows that a knowledge of the surface tension of the clean solid metal could not be used to facilitate the study of spreading when it is in contact with different liquid metals.

ACKNOWLEDGMENTS

The authors wish to express their thanks to the Director and Council of the British Non-Ferrous Metals Research Association for permission to publish this paper, which was released to members of the Association in September 1949.

REFERENCES

- BIRCUMSHAW, L. L., 1934, *Phil. Mag.*, **17**, 181.
- CHALMERS, B., KING, R., and SHUTTLEWORTH, R., 1948, *Proc. Roy. Soc. A*, **193**, 465.
- GURNEY, C., 1949, *Proc. Phys. Soc. A*, **62**, 639.
- HUANG, K., and WYLLIE, G., 1949, *Proc. Phys. Soc. A*, **62**, 180.
- SAUERWALD, F., and DRATH, G., 1927, *Z. anorg. Chem.*, **162**, 301.
- SHUTTLEWORTH, R., and BAILEY, G. L. J., 1948, *Discussion, Faraday Soc.*, No. 3.
- SMITH, C. S., 1948, *Amer. Inst. Min. Met. Engrs., Tech. Pub. No.* 2387.
- UDIN, H., SCHALER, A. J., and WULF, J., 1949, *J. Met.*, **1** (2), 186.
- WENZEL, R. N., 1936, *Ind. Eng. Chem.*, **28**, 988.
- WILHELMY, L., 1863, *Ann. Phys., Lpz.*, **119**, 177.

The Construction of Interference Filters for the Transmission of Light of Specified Wavelengths

By K. M. GREENLAND AND C. BILLINGTON

British Scientific Instrument Research Association, Chislehurst, Kent

MS. received 16th August 1949, and in final form 28th December 1949

ABSTRACT. Optical interference filters are composed of thin films deposited on a glass base by volatilization in a high vacuum. The narrowness of the transmission bands characteristic of these filters demands a high degree of accuracy in the regulation of the thickness of the films if maximum transmission is to occur at a specified wavelength. A method of observing the optical thickness of transparent films during deposition is described and it is concluded that its accuracy is sufficient, for instance, to enable the centre of a fourth-order transmission band to be placed within 20 Å. of any wavelength in the visible spectrum.

§1. INTRODUCTION

THE interference filter is composed of two partially reflecting metal films separated by a film of transparent material (Geffcken 1942, 1948, 1949). A full mathematical and experimental analysis of their properties is given in two papers by Hadley and Dennison (1947, 1948). In the construction of an interference filter intended to transmit light of a particular wavelength, it is essential that the thickness of the transparent film should be adjusted as accurately as possible to the value determined by that wavelength. The films are deposited on a glass base by slowly volatilizing the materials in a high vacuum; methods of control of film thickness may therefore depend either on the evaporation of a predetermined amount of material (measured by weight or by time and temperature of evaporation) or on an observed indication of the film thickness as the film grows.

Methods of the first kind rely on a degree of constancy in the conditions of volatilization which is in practice almost impossible to obtain. We have, therefore, devised an optical system by means of which the increasing optical thickness of the film in the vacuum chamber may be estimated visually with an accuracy sufficient to place the transmission band of the filter within about 20 Å. of the required wavelength.

The filters used in the work on which this paper is based were made of two partially reflecting silver films separated by a layer of zinc sulphide. It will be convenient to refer to silver and ZnS films, but the use of other materials is not precluded. It may be mentioned, in passing, that films of magnesium fluoride, a material commonly used for interference filters, were found to behave in an erratic manner after deposition; their optical thickness changed irregularly when air was admitted to the vacuum chamber and when a cover-glass was cemented to the filter. Zinc sulphide is not subject to these irregularities.

§2. PRINCIPLE OF THE METHOD

The transmission interference filter is also a highly reflecting mirror, but the spectrum of a white-light source reflected by the filter exhibits, in general, dark bands. These absorption bands are due to interference in the wave system set up between the silver films, and in a transmission filter they are approximately

complementary to the transmission bands. Owing to the high reflectivity of the silver films, the absorption bands, like the transmission bands, are very sharply defined.

The transmission bands cannot, of course, be observed during the critical stage of construction, that is, while the ZnS is being deposited, because they do not appear until the second silver layer has been formed. Absorption bands, on the contrary, are observable at this stage because it is possible to direct the incident beam so that total internal reflection at the growing ZnS surface may perform the function of the absent silver film. If, then, the relation between the wavelengths of the absorption bands and transmission bands is known, the wavelengths of one or more of the absorption bands may be used as an accurate measure for the evaporation of the ZnS.

The essence of the method is the achievement of total internal reflection at the growing ZnS surface; without this the absorption bands are so broad that accurate estimation of their mean wavelength is impossible. This condition is realized by introducing the 'monitor' beam into the filter base through a right-angled prism, so that the angle of incidence on the filter itself is 45° .

For a thickness t of ZnS film, absorption bands will occur in the monitor beam with maximum absorption at wavelengths λ_a satisfying the condition

$$(2\pi/\lambda_a)nt \cos \theta + \phi = m\pi, \quad \dots\dots(1)$$

where n is the refractive index of the ZnS film, θ is the angle of refraction in the ZnS, ϕ represents phase shifts due to reflection at the ZnS boundaries and transmission through the silver film, and m is the order of interference. The transmission bands of the complete filter, observed at normal incidence, will be centred on wavelengths λ_0 given by the condition

$$(2\pi/\lambda_0)nt + \phi_0 = m\pi, \quad \dots\dots(2)$$

where ϕ_0 represents the phase shift at the ZnS-silver boundaries. If the values of ϕ and ϕ_0 were known with sufficient accuracy, a filter having a transmission band centred at any desired wavelength λ_0 could be constructed by calculating from equations (1) and (2) the wavelength λ_a of the appropriate absorption band. As it is, these values for phase shifts are not yet known; the procedure is, therefore, to construct a number of filters for which both λ_a and λ_0 are measured and so to derive an empirical relationship between these two wavelengths. This does not require a knowledge of the phase shifts, but it is necessary that, for a given wavelength, they should not vary from one filter to another.

§3. CONTROL OF PHASE SHIFTS

As far as is possible we ensure that there is no variation in the experimental conditions, such as purity of materials, degree of vacuum, rate of evaporation and temperature of filter base; any of these factors may affect the quality of the silver films and hence alter the phase shifts. The only variables affecting the phase shifts which then remain to be controlled are the transmissivities of the silver films. The light in the monitoring beam must also be polarized in a suitable direction.

(i) *Transmissivity of the Silver Films*

The transmissivities of the silver films are set within fairly close limits by whatever transmission factor is required for the filter. For film transmissivities giving a filter transmission factor of 25% (a typical value) or less the value of the

reflection phase shift is not very sensitive to small changes in transmissivity, so that control of transmissivity is not such a critical matter as control of the ZnS film thickness. The easiest way to obtain a measure of control over the transmissivity of the silver films is to evaporate completely a fixed weight of silver for each film. This was, in fact, the method finally adopted for the present investigation, and results have shown that a sufficient uniformity of transmissivity is usually, but not always, obtained. It is essential to ensure that the silver vapourizes smoothly without spitting molten metal.

An optical method was devised by one of us (C.B.) for matching the optical density of the silver film being deposited with that of a 'standard' film of the required density. This method was tried but abandoned owing to the difficulty of keeping the optical system, which was entirely within the vacuum chamber, free from deposits of ZnS, which disturb the essential optical 'balance' of the system.

(ii) *Polarization of the Monitor Beam*

The values of the phase shifts due to reflection of the monitor beam at the two boundaries of the ZnS film depend on the direction of polarization of the light relative to the plane of the filter. If the beam is 'unpolarized', the absorption bands will occur in pairs (which we shall call the *p*-absorption and *s*-absorption bands) due to the differing phase shifts for the *p*- and *s*-components. The bands will then be ill-defined through lack of contrast, because each band is seen against a bright background of light polarized at right angles to that being absorbed. It is therefore necessary to polarize the beam in either the 'parallel' or 'perpendicular' plane relative to the filter in order to remove this unwanted light. It is in practice convenient to be able to change the plane of polarization since, for filters of low order for certain wavelength regions, one or other of the absorption bands falls near or outside the limit of spectral visibility. There is, in fact, a small region for which filters of the first order cannot be monitored directly (this difficulty might be overcome by using a separate monitor prism and placing the filter blank at a greater distance from the evaporator than the prism; a special calibration would, of course, be necessary).

§ 4. OPTICAL SYSTEM FOR THE MONITOR BEAM

The optical system for observing the absorption bands consists of an illuminating beam and a spectrometer to analyse the light reflected from the films deposited on the filter base. The glass filter base is supported over a hole in a horizontal shield near the top of the vacuum bell-jar, vertically above the silver and ZnS evaporators. It is illuminated by light from a Pointolite lamp incident at an angle of 45° on the upper side of the filter base. In order that the angle of incidence at the ZnS-vacuum boundary may be high enough for total internal reflection a right-angled prism is placed with its hypotenuse face on the top surface of the filter base, with a film of high-vacuum oil between the surfaces to ensure optical continuity. The reflected beam is directed by a mirror and lens on to the slit of a plane grating spectrometer adjusted for the diffraction spectrum of first order. It is essential for good definition that the spectrometer slit should be at the principal focus of the lens. In front of the spectrometer slit is a rotatable 'Polaroid' polarizer.

§ 5. EXPERIMENTAL PROCEDURE

For the purpose of calibrating the instrument—that is, of determining the relationship between λ_a and λ_0 —a number of filters are made with various thicknesses of ZnS film. The wavelength of the centre of either the *p*- or *s*-absorption band, λ_a , is read after the ZnS evaporation; the filter is then completed by evaporation of the second silver layer and removed from the vacuum chamber. A cover-glass is cemented on to the exposed silver surface and the wavelength of the centre of the transmission band, λ_0 , measured at normal incidence.

§ 6. DISCUSSION OF RESULTS

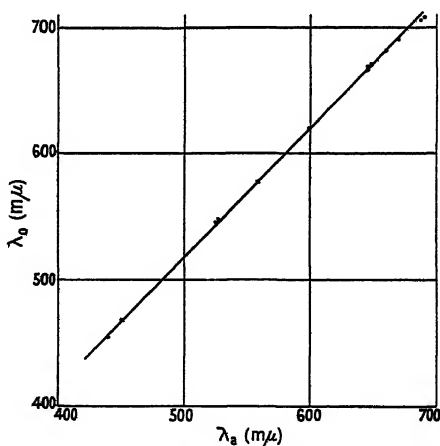
We find that the relation between the experimental values of λ_a and λ_0 may generally be expressed in the form

$$\lambda_a = A\lambda_0 + B. \quad \dots\dots (3)$$

Values of *A* and *B* for various orders are given in the Table. The figures relate to wavelengths measured in millimicrons and to λ_a measured for the *s*-absorption band.

Order	3	4	5
<i>A</i>	1.00	0.98	0.98
<i>B</i>	−6	−5	−15

The experimental results used in calculating the values of *A* and *B* for the fourth order are reproduced graphically in the Figure. These results were obtained from a series of 24 filters; four filters were rejected because of unevenness in



Calibration for fourth order.

the ZnS layer and the measurements from all the remainder were plotted (coincidence of values on the graph reduces the number of points to 13). The experimental values are indicated in the Figure by 'spots' having an equivalent diameter of approximately 20 Å. It can be seen that, excepting two red filters, all the filters measured had values of λ_0 lying well within 20 Å. of those satisfying the linear relationship of equation (3). Seven filters of this series were made with the monitor band wavelength λ_a adjusted to 526 mμ: the filters all had values of λ_0 lying between 544 and 546 mμ. It may therefore be said that control of λ_0 is possible within 20 Å. for the fourth order.

The high concordance of results for the fourth order has, in our experience so far, not been quite so well maintained at lower orders; discrepant filters occur either singly or in a batch of successive evaporations. This may be the result of changes in the value of phase shift on reflection at the silver-ZnS boundaries, due to variations in the thickness and structure of the silver films. Experiments now in progress show that although the evaporation of a fixed weight of silver gives a reasonably consistent transmissivity in the majority of evaporations, there can be a very large deviation from the normal value.

§7. CONCLUSION

Determination of the relationship between the mean wavelength of an absorption band observable during the deposition of the ZnS *in vacuo* and the mean wavelength of the corresponding transmission band of the complete interference filter has made it possible to place the peak transmission of the filter within 20 Å. of a predetermined wavelength. Greater discrepancies which sometimes occur may be due to an insufficient degree of control over the deposition of the silver films.

ACKNOWLEDGMENT

The authors desire to express their thanks to the Director and Council of the British Scientific Instrument Research Association for permission to publish this paper.

REFERENCES

- GEFFCKEN, W., 1942, German Patent No. 716,153; 1948, *Angew. Chem.*, A60, 1; 1949, *Ceram. Abstr.*, 32, 87.
HADLEY, L. N., and DENNISON, D. M., 1947, *J. Opt. Soc. Amer.*, 37, 451; 1948, *Ibid.*, 38, 483.

Note on the Brightness Profile and Photometric Contrast of a Test-object having small Angular Dimensions and Silhouetted against the Twilight Sky *

By D. R. BARBER

Norman Lockyer Observatory, of the University College of the South-west,
Sidmouth, South Devon.

*Communicated by S. C. Goddard; MS. received 18th July 1949, and in amended form
6th December 1949.*

ABSTRACT. Microphotometry of photographs of distant objects with small angular subtense has shown that inaccurate results are obtained if the apparent contrast is calculated on the assumption of uniform sky brightness. No evidence was found of any edge effect due to inward small angle scatter.

§ 1. INTRODUCTION

IN the course of an investigation of the variation of brightness contrast of the sea-horizon an opportunity was taken to test the theoretical conclusions of H. Koschmieder (1924) regarding the physical basis of aerial perspective.

A long-focus camera was used to photograph a series of distant test-objects of different albedos at a range of 4,850 yards. Each exposure was calibrated by means of a supplementary exposure to a step-tablet of known brightness range, and the set of negatives was analysed by the aid of high-magnification ($50\times$) microphotometer tracings of the photographic image of the test-object. Auxiliary exposures to a calibrated light source over a range of 2,000 yards were made concurrently with the main series in order to measure photographically the atmospheric transmission at the time of each observation.

Two separate objects were available for the experiments. One of these, marked T in Figure 1 (*a*) (Plate)†, had an angular subtense of about 3' of arc as viewed from the camera site, and an albedo of 0.09, corresponding to a photometric contrast of 0.91. The second, marked H in the figure, had been specially painted so as to provide a composite object of known differential contrast: the panels, a, b, c (see Figure 1 (*b*)) had albedos of 0.80, 0.047 and 0.86 respectively corresponding to photometric contrasts, relative to the dark panel b, of 0.94 (*a*) and 0.95 (*c*).

The photographs used for the subsequent analysis were taken either in early twilight or in moonlight, under varying conditions of cloud-cover and visibility. Examples of the test-object profiles, plotted on an intensity scale from the microphotometer tracings, are reproduced in Figures 2 (*a*) and 2 (*b*). Brightness values are in c/ft^2 and the broken curve indicates in each diagram the measured profile when corrected for the limited resolution of the photographic record. The 'peak' of this profile is plotted at a level of intensity computed from the known albedo of the test-object, measured at close range, and the intensity of the sky background. In computing contrast values from intensity

* This work was carried out while with the Kodak Research Laboratories and is issued as Communication No. 1262H from these.

† Opposite n. 368.

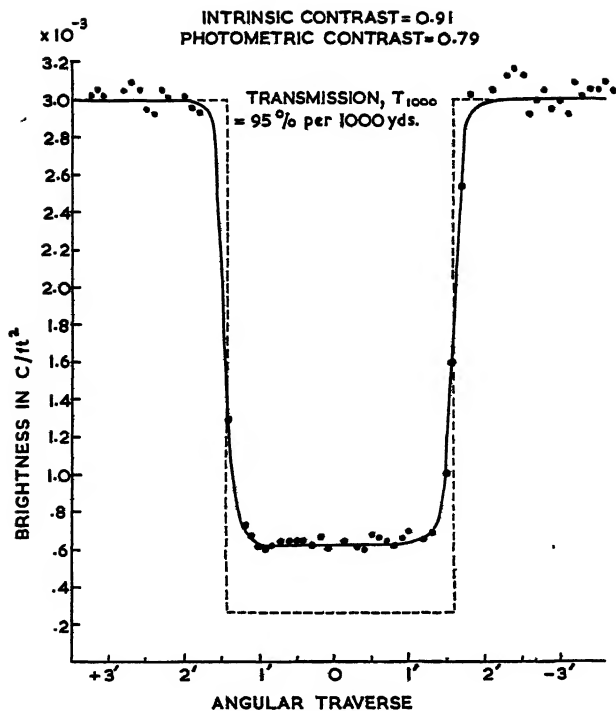


Figure 2 (a). Intensity profiles: — photometric, - - - - computed from albedo, of low-albedo test-object. The shape of the photometric profile results from the limited resolution of the original photographic record.

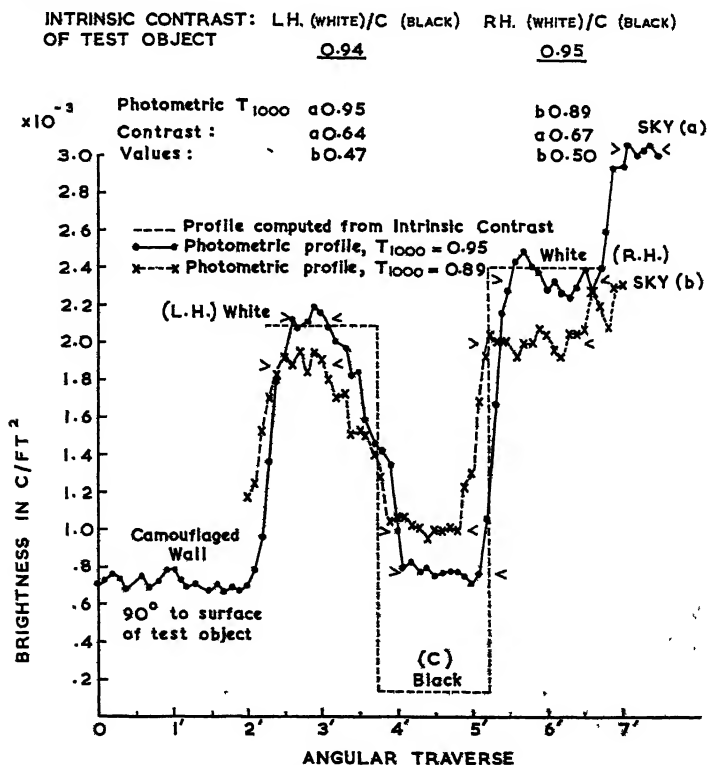


Figure 2 (b). Intensity profiles of composite test-object: — photometric, - - - - computed from albedo measures. Plotted to same scale.

measures the expression, $C = (B - B')/B$ is used, where B and B' are brightness-values for surround and test-object respectively; when the object appears lighter than its surround the contrast is taken to be negative.

§2. THEORETICAL TREATMENT

Considering now Koschmieder's argument, two conclusions emerge.

Firstly, the apparent brightness B_R of a test-object, albedo R , as observed from a distance l against the uniform background of horizon sky, brightness B_H , will be given by the expression,

$$B_R = RB_w e^{-\sigma l} + (1 - e^{-\sigma l}) B_H \quad \dots \dots (1)$$

where σ is the generalized extinction coefficient of the atmosphere at the time of observation, and B_w is the intrinsic brightness of a perfectly white object ($R = 1$).

On the assumption that the sky has uniform brightness in all azimuths, and that the brightness of the ground is zero, the value of B_w may be taken as $B_H/2$ so that the expression for the observed brightness of the test-object now becomes.

$$B_R = R(B_H/2)e^{-\sigma l} + (1 - e^{-\sigma l}) B_H, \text{ or } B_R = B_H[(R/2)e^{-\sigma l} + (1 - e^{-\sigma l})] \dots \dots (2)$$

Equation (2) may now be used to obtain the apparent contrast of the test-object, C_R where $C_R = (B_H - B_R)/B_H = 1 - [(R/2)e^{-\sigma l} + 1 - e^{-\sigma l}]$ from (2) or

$$C_R = e^{-\sigma l}(1 - R/2). \quad \dots \dots (3)$$

This equation, among others, has been used to compute the apparent contrast of a test-object observed under different conditions of atmospheric transmission; and, for the purpose, values of the extinction coefficient σ have been calculated from the measured values of T_{1000} , the transmission per 1,000 yards. They appear in column (2) of Table 1.

Table 1

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	
R	σ $\times 10^{-4}$	B_s (c/ft ²)	C_{ob}	C_c			$(C_c - C_{ob})/C_{ob}$			
				$n=0$	$n=1$	$n=2$	$n=0$	$n=1$	$n=2$	
0.05	1.20	Twilight	2.0	0.56	0.55	0.54	0.53	-0.018	-0.036	-0.054
0.80	1.20	„	„	0.19	0.32	0.21	0.10	+0.684	+0.105	-0.474
0.86	1.20	„	„	0.14	0.30	0.18	0.06	+1.142	+0.286	-0.571
0.09	0.81	Moonlight	8.0	0.63	0.64	0.62	0.61	+0.016	-0.016	-0.032
0.09	0.52	Twilight	12.0	0.72	0.75	0.72	0.70	+0.042	+0.000	-0.028
0.05	0.52	„	3.0	0.75	0.76	0.74	0.73	+0.013	-0.013	-0.026
0.80	0.52	„	„	0.33	0.43	0.28	0.13	+0.303	-0.152	-0.606
0.86	0.52	„	„	0.23	0.40	0.25	0.09	+0.738	+0.087	-0.608
0.09	0.23	„	„	0.79	0.82	0.80	0.78	+0.038	+0.013	-0.013

Secondly, it is shown by Koschmieder that both the profile, and the apparent contrast of a distant object of very small dimensions viz. less than about 20' of arc, should be influenced by inward small-angle scatter of light from the surrounding sky. Lohle (1929) has attempted a quantitative estimate of the 'edge effect' and finds that, for an object of 5' angular radius, this inward scatter will amount to 10% of the sky brightness. Such an effect, if present, will diffuse the boundary of the object (ground-glass effect), and thus distort its angular dimensions, in addition to decreasing the apparent contrast, as computed from equation (3) above.

There remains some doubt as to the reality of this inward scatter. Recently Fry, Bridgman and Ellerbrock (1947) have made a detailed theoretical analysis of both effects for a low-albedo test-object, surrounded on all sides by a uniformly bright sky. They conclude that, for objects of angular width 0.01 radian, both 'edge' and 'ground-glass' effects are negligibly small. Thus the boundary profile will, for all practical purposes, be undistorted. In these circumstances, the contrast, as viewed from a distance, may be measured satisfactorily by means of the intensity drop at the boundary of object and sky-surround. The experimental results to be presented later support these conclusions.

Koschmieder further emphasizes the influence of the threshold contrast value for the observer on the observed contrast of a distant object. In daylight this lies between 0.01 and 0.02 for the average eye, when observing objects of angular dimensions greater than about 2 degrees. The value increases rapidly with decreasing angular dimensions: for an angular subtense of 0.2 degree the threshold contrast is increased five-fold. However, this particular complication will not enter into the present investigation so long as the dimensions of the object remain such that the image on the photographic record is adequately resolved. Moreover, it should be emphasized that the test-objects used in the present work have dimensions much smaller than those referred to above. It is therefore of considerable interest to compute (i) the computed and measured contrast values for various atmospheric transmissions, and (ii) the profiles obtained through varying amounts of haze or mist with those recorded under atmospheric conditions.

Koschmieder's assumption, in common with other early investigators. (Föitzik 1932, Wright 1939, Middleton 1941), of a uniformly illuminated hemisphere of sky surrounding the test-object, and negligible ground reflectance in the vicinity of the object, may lead to appreciable errors if equation (3) is used to compute the apparent contrast values under very different atmospheric conditions; this is particularly so when the object considered has a relatively high albedo. This complication has been stressed by Moon and Spencer (1942) who show that the equation expressing the brightness, B_{ξ} , of an element of sky at a zenith angle ξ may be written

$$B_{\xi} = B_h(1 + 2 \cos \xi)$$

where B_h is the brightness of the horizon sky at $\xi = \pi/2$. This gives $B_{\text{zenith}} = 3B_{\text{horizon}}$, a value in line with that observed by Kimball and Hind (1921, 1922). More recently Middleton (1947) has extended the theoretical treatment by generalizing the above equation in the form,

$$B_{\xi} = B_h(1 + n \cos \xi)$$

where n is an integer, the remaining quantities being defined as before.

From this relation it follows that the brightness, B_w' , of a vertical white surface illuminated by skylight alone will no longer be given by the quantity $B_h/2$, but by the expression,

$$B_w' = B_h(\frac{1}{2} + 2n/3\pi).$$

Further, if R' is the albedo of the ground in the immediate vicinity of the test-object, the resultant brightness, B_w'' , due to ground reflectance, will be

$$B_w'' = R'B_h(\frac{1}{2} + n/3).$$

Thus the observed brightness of the test-object, due to the combined effects of sky and ground illumination, may be expressed as

$$B_w = B_w' + B_w'' = B_h[\frac{1}{2}(1 + R') + (n/3)(2/\pi + R')]$$

and, if the white object is replaced by one of albedo R ,

$$B_R = B_h[(R/2)(1 + R') + (nR/3)(2/\pi + R')].$$

Hence the apparent contrast, C_R' , of a grey object observed through an horizontal layer of atmosphere, thickness L and transmission coefficient σ , is given by the equation,

$$C_R' = [1 - \{(R/2)(1 + R') + (nR/3)(2/\pi + R')\}]e^{-\sigma L}. \quad \dots \dots (5)$$

Equation (5) replaces (3) and enables a correction to be made for the combined effects of a non-uniform sky illumination, and of ground reflectance.

Table 1 illustrates the magnitude of the error which may result if the simple equation is used to compute the observed contrast for objects of high and low albedo values, when measured in a horizontal direction over a range of 4,840 yards, under varying atmospheric conditions and silhouetted against the twilight sky. Values of the transmission coefficient σ ranged from 1.20×10^{-4} to 0.23×10^{-4} . A value of $R' = 0.09$ has been taken for the average ground reflectance* around the test-object. Column (1) of the Table gives the true albedo of each test-object observed, and column (2) the value of the generalized extinction coefficient at the time of observation. In column (3) there are entered particulars of the atmospheric conditions, and the observed brightness of the sky. Column (4) contains the experimentally observed contrast value, whilst columns (5), (6) and (7) refer to values of contrast computed by the aid of equation (5) taking $R' = 0$, $n = 0$ (column 5), and $R' = 0.09$, $n = 1, 2$ (columns 6 and 7). Finally, in columns (8), (9) and (10) there are given differences between computed and observed contrast utilizing the data of columns (4) to (7). In each instance, the difference is expressed as a fraction of the observed quantity.

§ 3. DISCUSSION

It will be seen that the contrast values computed for low-albedo objects agree well with those experimentally determined, so that the error involved in using the simplified equation (3) will, in general, be small. However, for high-albedo objects the error is in all cases appreciable. Indeed the computed value may be in error by as much as 100% of the observed contrast, as is illustrated by the case of the light-toned object ($R = 0.86$) viewed against the twilight sky. In this particular instance, the contrast difference (computed minus observed) amounted to 0.16 but fell to 0.04 when account was taken of non-uniform sky and ground reflectance. For all the observations listed in Table 1 the object measured was illuminated by light received from a cloudless sky. It was found that, for this condition the computed values fitted observations best when equation (5) was entered with $n = 1$. Table 2 summarizes the contrast data for the low-contrast object ($R = 0.09$) illuminated by a totally overcast twilight sky. Three values of C_c are found in adjacent columns: the first of these computed values is based essentially on equation (3) where $n = 0$, $R' = 0$, the remainder on equation (5) for $R' = 0.09$, $n = 1$, $n = 2$ respectively. Under an overcast sky the

* The site of the test-object was at a cliff edge on rough grassland with red sandstone subsoil and coarse wild herbage on the surface.

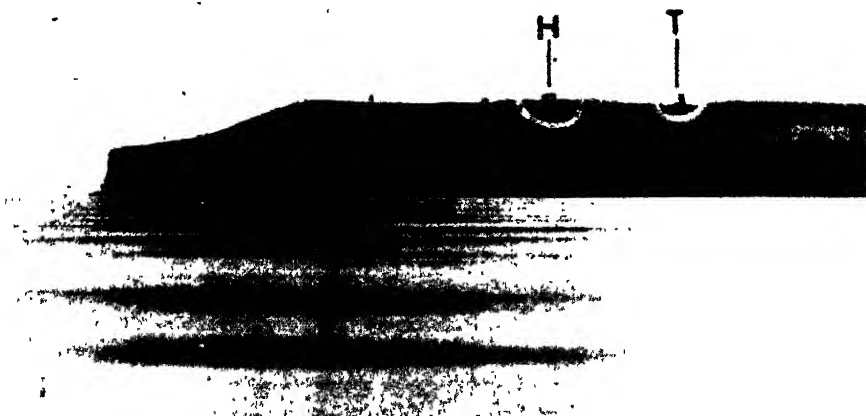


Figure 1 (a).

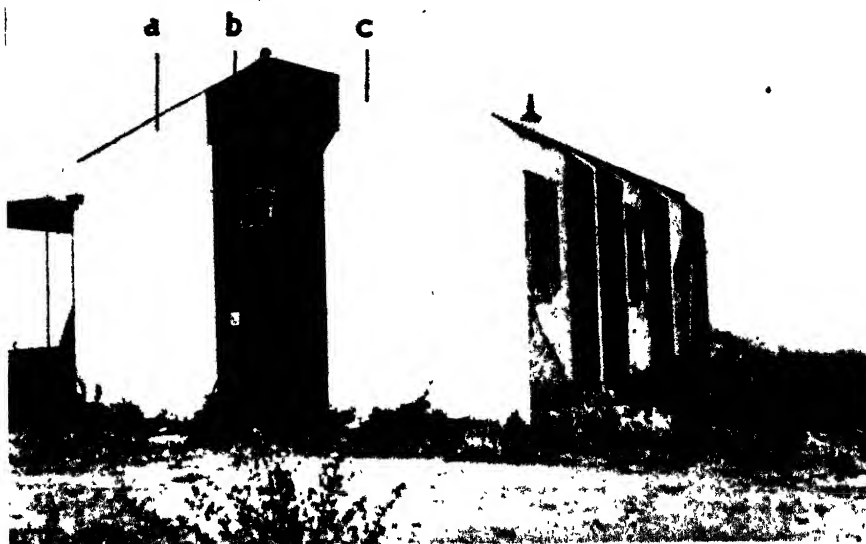


Figure 1 (b).

Figure 1. The test-objects (H and T) as seen from the camera site (a), and close-up view of composite test-object, H (b).

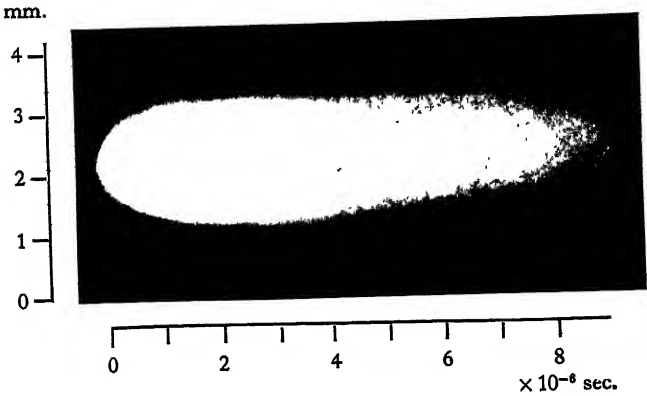


Figure 1.

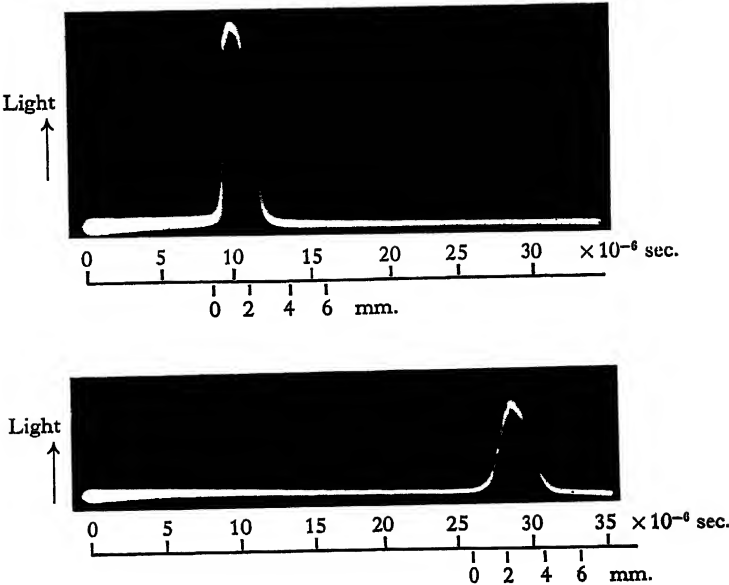


Figure 2.

computed contrast, using $n=2$, is found to fit experimental results rather better than that obtained from the use of equation (3) with $n=0$, or equation (5) with $n=1$.

It is evident that the simplified theoretical treatment based on a uniformly bright sky from horizon to zenith fails to yield results concordant with observation. By using Middleton's modification it is possible, however, to correct satisfactorily both for the sky effect and ground reflectance.

One other conclusion has emerged from a detailed study of the high magnification tracings of the brightness profiles of the low albedo test-objects. It is

Table 2

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	
R	σ $\times 10^{-4}$	B_s (c/ft ²)	C_{ob}	$n=0$	C_c $n=1$	$n=2$	$n=0$	$(C_c - C_{ob})/C_{ob}$ $n=1$	$n=2$	
0.09	3.55	Twilight	30.0	0.16	0.17	0.17	0.16	+0.063	+0.063	+0.000
0.09	1.55	"	4.0	0.42	0.46	0.45	0.43	+0.095	+0.071	+0.024

found that at high transmission values the profile, when properly corrected for the lack of resolution of the photographic lens-film combination employed in taking the original record, has a contour approximating very closely to that expected for a 'knife-edge' boundary. Further, the observed variations in the width of the object derived from the lower transmission records, made under conditions of light mist, or haze, exhibit no evidence of any 'edge-effect' due to inward light scatter, even when the object viewed has an angular width so small as 3 minutes of arc. These results support the theoretical arguments of Middleton (1942) and of Fry, Bridgman and Ellerbrock (1947).

ACKNOWLEDGMENT

The experimental work forming the basis of the present paper was carried out as part of a general programme of marine photometry undertaken by the Kodak Research Laboratories under contract with the Admiralty, to whom acknowledgment is made for permission to publish this paper.

REFERENCES

- FÖITZIK, L., 1932, *Met. Z.*, **49**, 134.
 FRY, G. A., BRIDGMAN, C. S., and ELLERBROCK, V. J., 1947, *J. Opt. Soc. Amer.*, **37**, 635.
 KIMBALL, H. H., and HIND, —, 1921-2, *Mon. Weath. Rev.*, Wash., **49**, 481; **50**, 616.
 KOSCHMIEDER, H., 1924, *Beitr. Z. Phys. freien Atm.*, **12**, 33-53; 171-81.
 LÖHLE, F., 1939, *Met. Z.*, **46**, 49-50.
 MIDDLETON, W. E. K., 1941, *Visibility in Meteorology* (2nd edn.) (Toronto: University Press); 1942, *J. Opt. Soc. Amer.*, **32**, 139; 1947, *Quart. J. R. Met. Soc.*, **73**, 456.
 MOON, P., and SPENCER, —, 1942, *Illum. Engng.*, N.Y., **37**, 707.
 WRIGHT, H. L., 1939, *Quart. J. R. Met. Soc.*, **65**, 411.

LETTERS TO THE EDITOR

The Measurement of Spark Channel Diameters

An attempt was made by Flowers (1943) to study the apparent optical diameters of spark channels, using a mechanical scanning camera. The results show that the optical diameter of the spark channel increases during the first few microseconds of its existence.

One of us (J.B.H.) has been extending this study in the course of work which will be published later. The technique used is essentially similar to that of Flowers, a rotating mirror being used to give the temporal resolution and a photographic emulsion for the recording. A record obtained in this way is shown in Figure 1 (see Plate opposite p. 368).

The light emission from a spark channel, and probably the distribution of light across the channel, varies rapidly. As little is known about the response of photographic emulsion subjected to the short exposures involved, it seemed desirable, as a check on the photographic method, to obtain spark-channel diameters using a different type of recording system. It was decided to use a photo-multiplier for this purpose.

The technique adopted involves a rotating mirror camera which scans the image of a spark across a slit, the length of which is parallel to the axis of the spark. The light passing through the slit is recorded by the photo-multiplier-amplifier-oscillograph unit described by Prime and Saxe (1949). The occurrence of the spark is synchronized with the position of the mirror, so that the image of the spark can be made to scan across the slit at any time relative to the initiation of the spark.

Figure 2 shows two typical records obtained in this way. The values of the spark-channel diameter are in good agreement with those obtained photographically. The records shown are slightly asymmetrical, principally because the Y-axis of the oscillograph is not quite normal to the X-axis. However, records taken during the earlier stages of the spark show a more pronounced asymmetry because of the change in the light distribution in the channel during the time of scan. For this reason no results have been obtained by this method during the first 5×10^{-6} second of the life of the spark channel.

The photographic method gives a continuous record of the variation with time of the diameter of a small portion of one particular spark channel, whereas the photo-multiplier technique gives only one cross section at a chosen time for each spark observed. However, the photomultiplier technique enables measurements to be made of the variation of light emission across the spark channel, and from these measurements it is possible to determine the radial distribution of light intensity in the channel. While the same result could be obtained photographically, there are difficulties involved because of the need for calibration of the photographic emulsion for the short exposures involved.

By using a higher mirror speed, a shorter and finer slit, and a greater amplification between the photo-multiplier and the oscillograph the new technique could be made to give a better temporal resolution and also to enable a greater range of light emission. In the examination of stable electrical discharges a slower scanning rate could be adopted, thus simplifying the amplifier design and enabling an ordinary oscillograph to be employed. The method may have applications in other fields.

This letter is published by permission of the Director of the Electrical Research Association.

Department of Electronics,
University of Liverpool.
January 1950.

* R. F. SAXE.
† J. B. HIGHAM.

FLOWERS, J. W., 1943, *Phys. Rev.*, **63**, 225.

PRIME, H. A., and SAXE, R. F., 1949, *J. Instn. Elect. Engrs.*, Part II, **96**, 662.

* Now at Armament Research Establishment, Fort Halstead, Sevenoaks, Kent.

† Now at Electrical Engineering Department, The University, Birmingham.

Crystal Diode and Triode Action in Lead Sulphide

Since the first publication on germanium crystal triodes by Bardeen and Brattain in 1948, it has been a matter of interest to investigate whether similar phenomena occur in other semiconductors. According to present theories of transistor action, the search should be directed towards materials which are *intrinsic* semiconductors at reasonably low temperatures and which have a high electron and hole mobility. These conditions are satisfied by PbS which also resembles Ge in other respects. For instance, both materials can occur as excess and deficit semiconductors, depending on the nature of the impurities present. Following this suggestion, selected samples of PbS were examined, using the usual arrangement as shown in Figure 1. These specimens were deficit semiconductors as inferred from

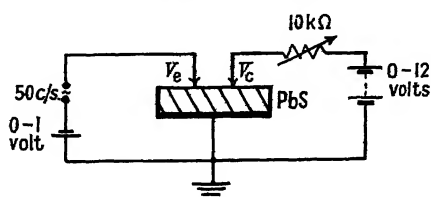


Figure 1. Schematic diagram of the circuit used for the detection of transistor action in lead sulphide.

(Distance between contact points is approx. 0.03 mm.)

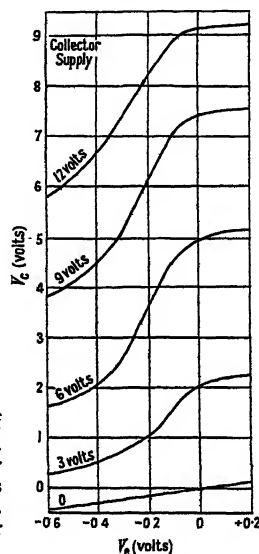


Figure 2. Dynamic characteristics of PbS transistor: relation between the emitter voltage V_e and collector voltage V_c for different values of the collector supply voltage. [Maximum gain = 13.] (Collector load = 10kΩ.)

the direction of rectification, which accounts for the polarity of the bias voltages shown. Transistor action was observed at some points of the material. A series of curves relating emitter voltage V_e and collector voltage V_c under dynamic conditions (load 10 kΩ) are given in Figure 2. It will be seen from the slopes that the voltage amplification is about 13 for 6 volts collector bias. Amplification factors as high as 25 have been recorded so far, though smaller values are more common on the existing samples. Power gains were in the neighbourhood of 4. Current gain has not so far been observed. The diode characteristics of emitter and collector are also of interest. Negative resistance may occur in both the forward and the reverse direction of current flow. Rectification ratios (at 1.5 volts) of 2000 : 1 have been obtained. All properties are appreciably affected by forming processes at the electrodes.

The above performance of the present PbS transistors, considered as circuit elements, is, of course, inferior to that of Ge transistors. It is, however, of interest to note that the transistor effect can occur in a solid of the PbS type. A full report on these investigations will be published at a later date.

The authors would like to thank Professor R. W. Ditchburn for placing research facilities at their disposal and Dr. H. K. Henisch for much helpful advice. (One of us (H.A.G.) is a member of the Royal Naval Scientific Service.)

Department of Physics,
The University,
Reading, Berks.
23rd February 1950.

H. A. GEBBIE.
P. C. BANBURY.
C. A. HOGARTH.

REVIEWS OF BOOKS

Principles of Physics.—Part III. *Optics*, by F. W. SEARS. Pp. 385. 3rd Edition. (Cambridge, Mass.: Addison Wesley Inc., 1948.) \$4.50.

The subject matter of this excellently produced book corresponds roughly to what used to be understood by 'pass degree standard'. The treatment, notably the mathematical part of it, is rather more elementary; on the other hand, the text gains considerably from the direct physical approach which this implies. Explanation is a difficult thing: to succeed demands of the expositor both clarity in his own mind and experience. It is clear that the author of the present work has full measure of both of these qualities.

The first four chapters deal with the propagation of light (including the determination of c) and the Gaussian properties of mirrors and lenses. The latter are considered from the point of view of wave theory. Chapter 5 introduces the subject of aberration, regrettably in the basis of ray theory.

It is in Chapter 6 that errors abound. This is perhaps understandable in a chapter on optical instruments—a subject in which many supposed specialists seem to be so lacking. It is suggested that the "aberrations of a simple magnifier may be greatly reduced (in the case of a simple Ramsden eyepiece) since four refracting surfaces are available". This is not the case. The additional two surfaces belong to a field lens, which increases most of the aberrations by a small amount and introduces very considerable field curvature. The important achromatized Ramsden is not mentioned. Again, on page 157, occurs the remarkable claim that "Most modern high speed short focal length lenses are modifications of the Zeiss 'Tessar' lens". In fact the 'Tessar' was simply an obvious modification of the well-known 'Cooke Triplet' lens due to the English designer H. D. Taylor. On page 144, Figure 6-15, it is the 'one' on the left, not the right, which is an oil immersion objective.

Chapters 7, 8 and 9 treat polarization, interference and diffraction respectively. The approach is very clear, but almost qualitative. Chapter 10 deals with the limit of resolution and somehow manages to avoid all reference to Abbe's theory of the microscope. The electron microscope is described, and (rather pointlessly) one and a half pages are devoted to prints of the 'soda-fountain' type of instrument. In the face of this, no mention is made of the very important topic of ultra-violet microscopy!

The remaining chapters deal conventionally, but adequately, with the elementary aspects of spectra, radiation, photometry and colour. Giving the whole gamut of terminology in these latter subjects will leave the student bewildered, but well-informed if he does stay the course.

There seems to be no attempt in the book to justify the addition of amplitudes or intensities according as the light is coherent or incoherent. That the intensity may be expressed by the square of the amplitude also seems to be stated rather than proved. One might have expected these two fundamental subjects to receive some discussion.

The standard of printing and the excellence of the diagrams should be singled out for special praise. The plates illustrating diffraction represent some beautiful experimenting. (It should be noted, that Figure 9-3 is described as Fresnel diffraction in the caption, but as Fraunhofer diffraction in the text, page 226. The text is correct.) Many of the diagrams revert to the practice of white-on-black, as an older generation will recall to have been used in Edser's *Light*.

The criticisms that have been given of this book should not detract from the general impression of considerable merit.

H. H. HOPKINS.

A Survey of General and Applied Rheology, by G. W. SCOTT BLAIR. Pp. xvi + 314. Second Edition. (London: Sir Isaac Pitman & Sons Ltd., 1949.) 40s.

Dr. Scott Blair is one of the pioneers of rheology in this country and has recently been elected president of the British Rheologists' Club. His survey of general and applied rheology was first published in 1943 and there now appears a second edition bringing the work up to date.

The book contains an historical and general introduction to rheology, followed by a wide survey of rheological phenomena. The survey is conducted under two main headings: (1) rheological phenomena and their measurement and (2) rheological interpretation and the evidence of psycho-physics, with the view to re-establishing "that association between our science and psychology which was so strongly emphasized by the early Indian rheologists". The book is addressed primarily to the practical rheologist and the physicist will not find it easy reading. For example, how many physicists would identify the ideal liquid of hydrodynamics from the description that "it resists pantathenic stresses, but without change of volume and has no deviator components in its stress tensor"? Presumably the practical rheologist understands this language. Any physicist contemplating an excursion into the field of rheology will, however, find certain sections of the book particularly useful. These are (1) the classification of methods of measurement where, under each method, there appears a list of materials to which the method has been applied, each material being followed by appropriate references; (2) the brief summaries indicating the scope of papers on certain subjects, e.g. viscosity of liquid mixtures, thixotropy, and dilatancy; and (3) the comprehensive bibliography. These three sections occupy some 70 pages of the book.

J. W. FOX.

CONTENTS FOR SECTION A

	PAGE
Prof. A. O. RANKINE. Experimental Studies in Thermal Convection (33rd Guthrie Lecture)	417
Dr. R. SHUTTLEWORTH. The Surface Tension of Solids	444
Mr. OM PARKASH and Mr. P. L. KAPUR. On the Temperature Dependence of Counter Characteristics in Self-quenching Geiger-Müller Counters	457
Dr. G. KORTUM and Dr. H. VERLEGER. The Pressure Broadening of the Vibration Rotation Band of the HCN Molecule at 10380A.	462
Dr. S. M. NAUDÉ and Dr. H. VERLEGER. The Vibration-Rotation Bands of the Hydrogen Halides HF, H ³⁵ Cl, H ³⁷ Cl, H ⁷⁹ Br, H ⁸¹ Br and H ¹²⁷ I	470
Dr. M. S. PATERSON. Calculation of the Correction for Instrumental Broadening in X-ray Diffraction Lines	477
Dr. B. BLEANEY and Dr. J. H. N. LOUBSER. The Inversion Spectra of NH ₃ , CH ₃ Cl and CH ₃ Br at High Pressures	483
Mr. W. M. GIBSON and Dr. L. L. GREEN. The Neutrons Emitted in the Bombardment of ⁶ Li by Deuterons	494
Dr. K. J. LE COUTEUR. The Relationship between the Size of Nuclear Disintegration Stars and the Mean Energy of the Emitted Particles	498
Dr. A. B. BHATIA. On the Scattering of Polarized Neutrons by Protons	502
Dr. L. MICHEL. Interaction between Four Half-Spin Particles and the Decay of the μ -Meson	514
Mr. F. D. S. BUTEMENT. Radioactive Dysprosium Isotopes	532
Letters to the Editor:	
Dr. J. V. JELLEY. The Radioactivity of Fluorine 20	538
Miss P. ROTHWELL and Mr. D. WEST. Proportional Counters in a Magnetic Field; an Investigation of the Nuclear Isomerism in ⁸⁰ Br	539
Miss P. ROTHWELL and Mr. D. WEST. The Calibration of Proportional Counters with X-rays produced by Radioactive Sources	541
Dr. R. L. F. BOYD. He ₂ ⁺ in the Helium Discharge	543
Dr. M. AFAF. Bands of ZrF	544
Contents for Section B	546
Abstracts for Section B	546

ABSTRACTS FOR SECTION A

The Surface Tension of Solids, by R. SHUTTLEWORTH.

ABSTRACT. A distinction is made between the surface Helmholtz free energy F , and the surface tension γ . The surface energy is the work necessary to form unit area of surface by a process of division: the surface tension is the tangential stress (force per unit length) in the surface layer; this stress must be balanced either by external forces or by volume stresses in the body.

The surface tension of a crystal face is related to the surface free energy by the relation

$$\gamma = F + A(dF/dA),$$

where A is the area of the surface. For a one-component liquid, surface free energy and tension are equal. For crystals the surface tension is not equal to the surface energy. The standard thermodynamic formulae of surface physics are reviewed, and it is found that the surface free energy appears in the expression for the equilibrium contact angle, and in the Kelvin expression for the excess vapour pressure of small drops, but that the surface tension appears in the expression for the difference in pressure between the two sides of a curved surface.

The surface tensions of inert-gas and alkali-halide crystals are calculated from expressions for their surface energies and are found to be negative. The surface tensions of homopolar crystals are zero if it is possible to neglect the interaction between atoms that are not nearest neighbours.

On the Temperature Dependence of Counter Characteristics in Self-quenching Geiger-Müller Counters, by OM PARKASH and P. L. KAPUR.

ABSTRACT. The effect of temperature on the counting rate-voltage characteristics of self-quenching Geiger-Müller counters with internal and external cathodes (in this paper denoted as counters Nos. 1 and 2) has been investigated within the temperatures ranging from 8° to 60° C.

It is found that the average counting rate remains independent of temperature (within the statistical error limits), and this constancy is better in the case of the external cathode counter than with the other. As counter No. 1 shows a greater increase in slope at higher temperatures, the rate appears to be greater at higher applied potentials. Further, it is noticed that the plateau decreases and disappears at lower temperatures in the case of counter No. 1, while counter No. 2 does not show this effect. This is partly explained by the formation of semiconducting paths between the central wire and the cathode, the discharges along which give rise to spurious counts in the case of counter No. 1, and the absence of these in the case of counter No. 2.

It is also observed that the slope of the plateau increases and the width decreases with rise of temperature in both cases, though the increase in slope is more marked with counter No. 1 than with counter No. 2. The increase in slope is probably due to the presence of a greater number of multiple discharges or spurious counts at higher temperatures, an important point which is under study.

The investigation also brings about the advantage of the external cathode in the construction and design of G-M counters, especially when reliable observations are desired with a counter under widely changing temperature conditions.

The Pressure Broadening of the Vibration Rotation Band of the HCN Molecule at 10380 Å., by G. KORTÜM AND H. VERLEGER.

ABSTRACT. The relation between the line-width of the 0-3 vibration-rotation band of HCN at 10380 Å. and the pressure of the HCN gas, as well as the pressure resulting from the addition of different extraneous gases, was investigated. In the P and R branches of the

band a relation was found between the half-line width and the rotational quantum number, J , which differs for the broadening due to self-pressure and the broadening resulting from pressure of extraneous gases. The results could be explained, according to London's theory, in the former case as an effect due to the combined action of rotational resonance alignment, and in the latter case to alignment alone. Measurements made with the addition of DCN suggest that the chain association of the hydrogen bridge bonds also has an effect on the half-width value.

The Vibration-Rotation Bands of the Hydrogen Halides HF, H³⁵Cl, H³⁷Cl, H⁷⁹Br, H⁸¹Br and H¹²⁷I, by S. M. NAUDÉ AND H. VERLEGER.

ABSTRACT. The 2-0, 3-0 and 4-0 bands of ¹H¹⁹F, the 4-0 and 5-0 bands of ¹H³⁵Cl and ¹H³⁷Cl, the 4-0 band of ¹H⁷⁹Br and ¹H⁸¹Br, and the 4-0 band of ¹H¹²⁷I were photographed under great dispersion in the infra-red. The molecular constants have been calculated and are summarized in a table.

Calculation of the Correction for Instrumental Broadening in X-ray Diffraction Lines, by M. S. PATERSON.

ABSTRACT. A new numerical procedure, based on relaxation methods, for calculating the correction for instrumental broadening in x-ray diffraction lines has been evolved. The corrected integral breadth and the true line profile are obtained without restrictive assumptions. The calculations for a practical example are given.

The Inversion Spectra of NH₃, CH₃Cl and CH₃Br at High Pressures, by B. BLEANEY AND J. H. N. LOUBSER.

ABSTRACT. The absorption due to the inversion spectrum of ammonia has been studied at pressures up to 6 atmospheres between wavelengths of 8 cm. and 8 mm. The resonant frequency is found to shift downward from 0.78 cm⁻¹ as the pressure rises, becoming substantially zero above 2 atmospheres. The line width rises less rapidly than the pressure in the transition region, but ultimately becomes proportional to the pressure with an effective 'collision diameter' of 7.7 Å. These effects are attributed to multiple collisions, which become dominant at the higher pressures; the resonant inversion is then destroyed by the molecular interaction.

In the methyl halides the inversion frequency is virtually zero owing to the low probability of tunnelling through the barrier, but the absorption line is spread out to the centimetre wave region by the collision process. The collision diameters are found to be 7.7 Å. and 9.0 Å. for methyl chloride and bromide respectively.

The Neutrons Emitted in the Bombardment of ⁶Li by Deuterons, by W. M. GIBSON AND L. L. GREEN.

ABSTRACT. The neutrons from deuteron bombardment of a separated ⁶Li target have been studied by the photographic plate method. The energy release for the reaction



has been shown to be 3.40 ± 0.05 mev., and evidence is given for the existence of a level in ⁷Be at 450 ± 60 kev. This is identified with the level at 430 kev. found in work on the reaction ¹⁰B(pα)⁷Be.

The Relationship between the Size of Nuclear Disintegration Stars and the Mean Energy of the Emitted Particles, by K. J. LE COUTEUR.

ABSTRACT. Nuclear evaporation theory predicts that the mean energy of the protons evaporated from disintegration stars should increase with star size. It is shown that the calculated rate of increase is in agreement with experimental results.

On the Scattering of Polarized Neutrons by Protons, by A. B. BHATIA.

ABSTRACT. The scattering and the consequent depolarization of a polarized beam of neutrons by an unpolarized proton gas is suggested as a means of obtaining information about the interaction potential between a neutron and a proton. It is found that the interaction, if it were spherically symmetrical, could be determined from a knowledge of the polarization and scattering differential cross sections. A similar analysis with non-central interaction has not yet been carried out. In this case, however, the polarization depends also on the azimuth ϕ of scattering. This dependence of polarization on ϕ is calculated for (a) very low energies, (b) very high energies. The maximum azimuthal variations for incident neutrons of energy 1 mev. and 100 mev. are about 1% and 50% respectively.

It is hoped that in future experimental techniques for the production and measurement of a polarized beam of fast neutrons will be developed, since these would give important information about the interaction potential between a neutron and a proton.

Interaction between Four Half-Spin Particles and the Decay of the μ -Meson, by L. MICHEL.

ABSTRACT. The general direct coupling between four fermions is studied. It is a linear combination of the five invariants used in β -decay theory. Altering the order of the particles in the Hamiltonian changes only the coefficients of the linear combination. The formalism of charge conjugation is used with the ordinary theory or the Majorana abbreviated theory for neutral particles. This is applied to the study of the decay of the μ -meson into an electron and two neutrinos.

Radioactive Dysprosium Isotopes, by F. D. S. BUTEMENT.

ABSTRACT. A new isotope of dysprosium, identified as ^{166}Dy , has been studied. It has been shown that it is formed by thermal neutron capture in the 140 minute isomer of ^{165}Dy , the capture cross section being about 1,200 barns. It decays with a half-life of 82 hours and is the parent of 27 hour ^{166}Ho . It emits beta-particles, the energy of which was found by absorption in aluminium to be 0.22 mev.; and gamma rays of less than 50 kev. energy.

Another new isotope, ^{159}Dy , has been made both by slow neutron irradiation of dysprosium and by deuteron bombardment of terbium. It decays by orbital electron capture, with a half-life greater than 50 days. It will be described in a later paper.

THE PROCEEDINGS OF THE PHYSICAL SOCIETY

Section B

VOL. 63, PART 6

1 June 1950

No. 366 B

The Behaviour of the Cathode Spot on an Undisturbed Liquid Surface of Low Work Function

By K. D. FROOME*

Department of Physics, Imperial College, London

MS. received 19th September 1949

ABSTRACT. The cathode spot of transient arcs using the liquid sodium-potassium alloy for cathode has been studied by means of the Kerr cell apparatus capable of taking a sequence of photographs. It is found that for a given current the spot takes, as in the case of mercury, the form of a line or broken line of total length roughly proportional to the current. This is also true for a spot in a magnetic field. Again as for mercury, if the current rises slowly the length of line increases proportionally to the current. If the rate of growth of current is greater than a value lying approximately between 10^8 and 2×10^8 amp/sec., then fresh spots form and spread out radially from the point of formation into thin semicircular or circular lines moving with a maximum radial velocity of about 10^4 cm/sec. The apparent current density of emission lies between 2.5 and 5×10^6 amp/cm², and consequently is higher than for mercury.

§1. INTRODUCTION

THE results described in this paper have been taken by means of the Kerr cell apparatus capable of taking a series of superimposed photographs which show a number of positions of the cathode spot as it moves over the cathode. The interval between each exposure can be varied from a fraction of a microsecond to many microseconds, and the number of exposures of the cathode spot taken during a discharge can be adjusted from one to as many as are required to build up a complete picture of the behaviour of the emitting area. The exposure can be varied independently of either interval or number. This apparatus has been described elsewhere (Froome 1948).

Observations are made by means of a low-power microscope arranged to look down on to the liquid cathode surface, the Kerr cell itself being placed just outside the microscope eyepiece. The actual optical arrangement used is shown in Figure 1 of a paper describing details for an undisturbed mercury cathode (Froome 1949); reference should be made to this paper for other allusions to mercury cathodes.

The transient discharges studied are obtained by discharging a single condenser, or a number of condensers, in the form of artificial lines, charged to approximately 100 volts through the tube and series impedance. The former is used when varying currents are required, and the artificial lines can be used to produce constant current pulses of 10 to 450 amperes.

* Now at the National Physical Laboratory, Teddington, Middlesex.

The discharge is termed an arc because the cathode spot emits a large number of electrons, and because there is a low overall potential drop. The term cathode spot is used to denote the whole emitting area on the cathode irrespective of its exact form.

The liquid alloy used has a constitution of approximately 25% sodium and 75% potassium, but almost identical results can be obtained with any other proportion provided it is liquid at room temperature. The alloy has been chosen for two reasons. Firstly, it is liquid at room temperature, and a liquid surface is desirable because it is reproducible; secondly, it has a low work function, whereas mercury has a high work function (4.5 volts). The curves given by Thomson and Thomson (1928) for photoelectric emission from the alloy indicate a work function of about 2 volts.

In order to use this alloy, the discharge tube must be sealed off. It was not found possible to purify the alloy by the technique of evaporation usual for the alkali metals, for this separated the constituents of the alloy and deposited potassium in an irremovable film over the discharge tube window. The following technique eventually proved entirely satisfactory.

The tube is made with a long thin glass pumping line attached. This line consists of a number of bulbs of about one inch diameter separated from each other by a narrow capillary neck. The end farthest from the discharge tube (and nearest the liquid air trap) is fitted with a cone joint device by means of which the previously prepared and consequently rather dirty alloy can be introduced into the vacuum line. The whole is out-gassed as thoroughly as possible and filled with argon to the operating pressure (0.1 to 0.01 mm.). Then the liquid alloy is tipped into the bulb farthest from the discharge tube and the whole sealed off from the pumps. The alloy is then passed from one bulb to the next through the narrow joining capillary tubes, and by the time it reaches the discharge tube it appears brilliantly clean. The discharge tube is then sealed off from the cumbersome remaining line of bulbs.

The type of discharge tube used is almost the same as that described earlier (Froome 1949, Figure 1), but no amalgamated copper foil is wrapped round the inside of the cathode neck. Consequently the alloy surface is always somewhat curved, and for this reason some of the photographs shown are in focus only in one or two successive exposure positions.

Finally, the cathode is cleaned further by running the tube for several hundred flashes at high peak current. This is absolutely essential if reproducible results are to be obtained. The cathode spot has a strong cleansing action upon the cathode.

In all the Kerr cell photographs reproduced the cathode spot has started from the bottom of the picture. Hence each photograph shows a number of positions of the emitting cathode area separated by the time interval stated, with time increasing as the cathode spot moves upwards and, in the case of rapidly rising currents, outwards as well. The discharge is initiated by means of a brief high tension pulse applied outside the discharge tube near the cathode. It is found that the cathode spot almost invariably forms on the liquid surface adjacent the glass wall of the tube nearest the position of application of the triggering pulse. In some of the photographs (notably those shown in Plate II) the curved surface of the glass cathode wall can be seen in black outline along the bottom of the picture.

§2. RESULTS WITH CONSTANT CURRENT PULSES

It is difficult with this alloy to obtain results from an undisturbed surface. The emitting area rapidly breaks up into a large number of minute separate spots. These 'unit spots' seem to form in self-made dents in the surface. The same behaviour is observed for mercury, but takes considerably longer.

Figure 1 (Plate I) shows a photograph with exposure 1 microsecond taken 4 microseconds from the start of a constant current discharge of 410 amperes. It is seen that even in this short time unit spots have formed, whereas the corresponding time for mercury would have been 20–30 microseconds or more.

Such observations suffer from the limitations imposed by these dents (Froome 1949); and much of the work for constant current pulses has therefore been undertaken with the cathode spot in a strong magnetic field, since this produces a movement of the cathode spot sufficiently rapid to inhibit unit spot formation.

For the cathode spot in a magnetic field we again find, as for mercury, a marked reluctance to travel with a velocity of more than 10^4 cm/sec. Also, if the velocity is much below this value, excessive formation of unit spots takes place, and these are left behind the main emitting area until they eventually extinguish.

Nevertheless, there is a marked similarity in behaviour to that for mercury. Figure 2 (a)–(d) (Plate I) shows sequences of exposures of the cathode spot for a constant current of 90 amperes at two magnetic field strengths and with two different exposures. The magnetic field is across the picture, the spot moving in exactly the opposite direction to the ponderomotive force in the emitted electrons. In Figure 2(a) the field is 900 gauss, in (b) it is 1,800 gauss. The exposure in each case is 0.1 microsecond, and the interval between exposures is 4.2 microseconds. At the higher field it is seen that the emitting area is in the form of a line similar to that for mercury, but even at this field strength, with the consequent high velocity of the spot, there is evidence of the formation of a few unit spots which have been left behind in the wake of the main emitting area. At the lower field the line has become extremely ragged, or broken, due to such spot formation, and thus it is difficult to assess the actual emitting area.

Figure 2(c) shows a sequence, with longer exposures (0.3 microsecond) but with the same interval between each, of another discharge again in a field of 900 gauss. The raggedness is more clearly seen—as are the numbers of unit spots. Figure 2(d) shows a spot in the higher field (1,800 gauss) with this longer exposure. The line-like structure of the main emitting area is again seen, but there is also evidence of unit spots which have broken off and been left behind. This behaviour is due to the dents such unit spots apparently make in the liquid surface. Once a dent has formed *the spot is not capable of rapid movement*; this applies to mercury also.

Figure 3 (a)–(c) (Plate I) shows multiple exposures of cathode spots carrying various currents in a magnetic field of 1,300 gauss. The exposure in each is 0.3 microsecond, and the interval between consecutive exposures is 4.2 microseconds; (a) shows the spot for a current of 14 amperes, (b) for 60 amperes, and (c) for 165 amperes. The emitting area is seen to be chiefly a line for the two lower currents, but for the highest it consists of a very ragged (broken) line plus a large number of unit spots in its wake. It is also seen that these spots have a life-time of the order of the interval between consecutive exposures, namely about 4 microseconds.

Figure 4 shows graphically a plot of the current against length of this emitting line. As for mercury, this curve is obtained from a large number of different results. In this case, however, only those results which show the best line-structure have been included, and a correction has been attempted for the probable effect of the residual unit spots.

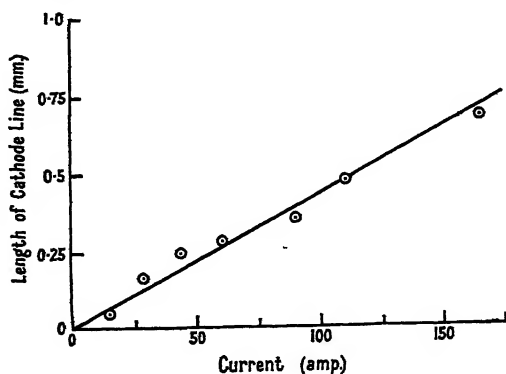


Figure 4.

As for mercury, we see that the length, l , of the emitting line is proportional to the current, i (amp.), where $l = Ki$ cm., and $K \simeq 4.1 \times 10^{-4}$ cm/amp $^{-1}$. This is under one-half the value obtained for mercury.

§ 3. RESULTS WITH VARYING CURRENTS

The results with varying currents show a strong similarity to the corresponding behaviour of mercury. As before, these results are obtained with the spot not in a magnetic field, in order to maintain simple conditions.

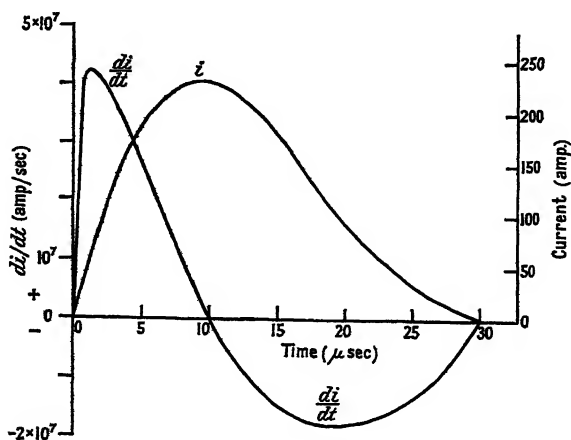


Figure 6.

Figure 5 (Plate II) shows the appearance of the cathode spot for the electrical conditions shown in Figure 6. It will be remembered that for mercury these conditions produced the formation of fresh spots. Figure 5 shows three exposures of 0.3 microsecond separated by 4.2 microseconds, the first being taken 2 microseconds from the initiation of the discharge. It is seen that only the first shows anything of a line-like structure, and that in the second and third slow-moving



Figure 1.

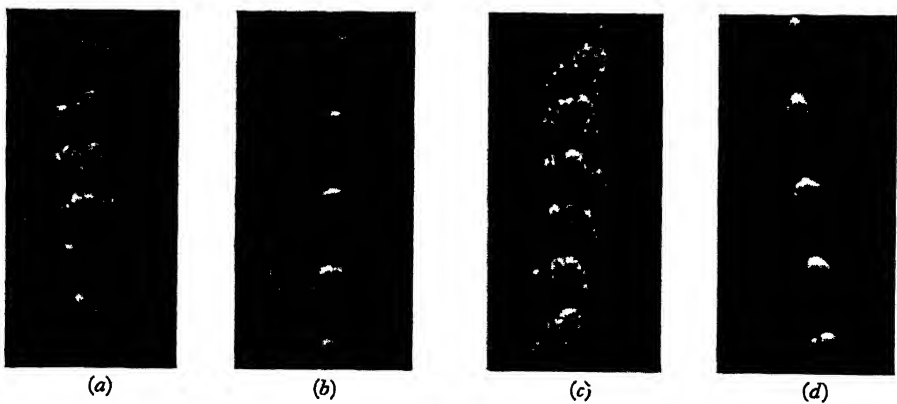


Figure 2.

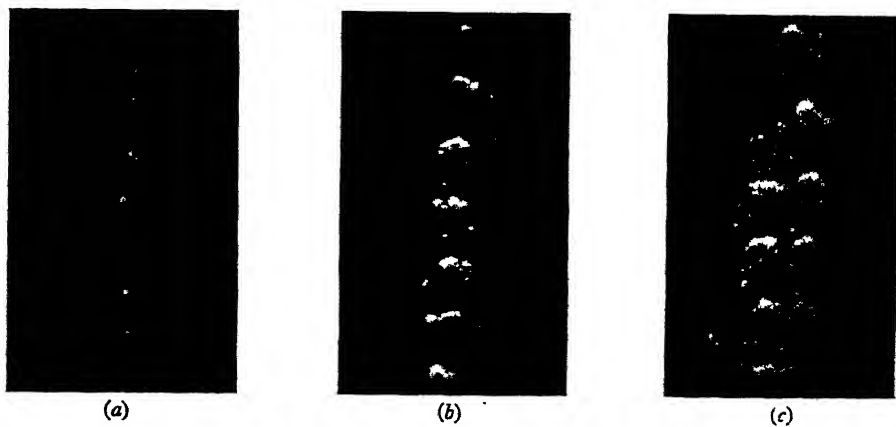


Figure 3.

PLATE I.



1 mm.

Figure 5.



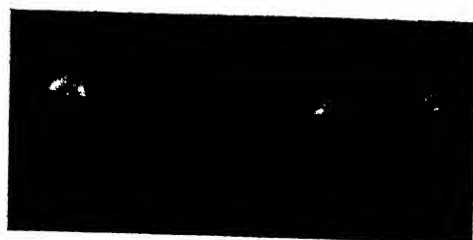
Figure 7.



Figure 9.



(a)



(b)

Figure 10.

PLATE II.

unit spots have formed; consequently in each of the last two exposures one can scarcely distinguish the positions of individual spots.

We have already seen that for a spot in a magnetic field the length of the emitting line for a given current is under one-half that of mercury. Hence, since the spot is reluctant to move much faster than 10^4 cm/sec., one can expect that at very high values of rate of growth of current the cathode spot will be forced to move in the form of a semicircular line at its maximum velocity, or to form new spots.

When the emitting area does take this form of a semicircular line moving outward with a maximum permitted velocity of 10^4 cm/sec., its maximum length L after t seconds will be given by $L = 10^4 \pi t$ cm., or, after t sec., the maximum current i emitted by such a line will be given by L/K where $K \simeq 4 \times 10^{-4}$. Hence the maximum current $i = 8 \times 10^7 t$, and therefore $di/dt = 8 \times 10^7$ amp/sec.

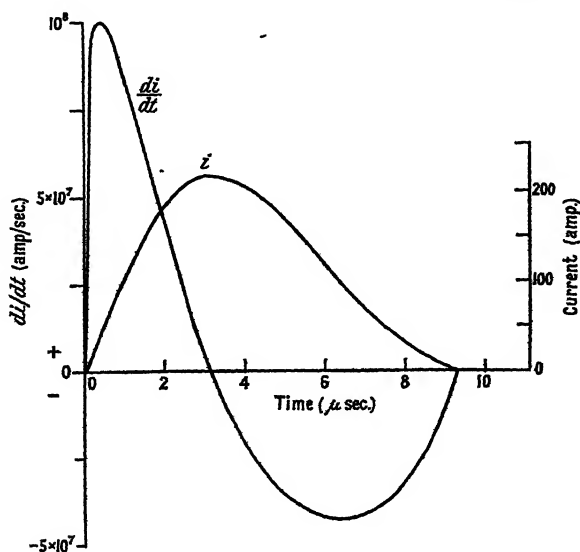


Figure 8.

represents the maximum rate of growth of current which we can expect to be maintained by a single semicircular spot. (The spot is almost invariably semicircular since it starts adjacent to the wall of the glass tube and can therefore spread out only into the tube. If the spot does form on the cathode surface away from the tube wall, then it can form a completely circular line, so that the maximum rate of growth of current sustainable by such a spot will be twice the above value.)

Figure 7 (Plate II) shows a sequence of four exposures (exposure 0.1 microsecond, interval 1.0 microsecond, the first exposure occurring just under a microsecond from the start of the arc) of the cathode spot obtained by discharging a $20 \mu\text{F.}$ condenser initially at 60 volts through the tube in series with a total inductance of about $0.3 \mu\text{H.}$ with negligible total circuit resistance. The peak value of rate of growth of current is 10^8 amp/sec. The cathode spot has formed as a single spot adjacent to the tube wall at the bottom of the picture, and then disintegrated into a semicircular line moving radially outward with a velocity of about 10^4 cm/sec. The characteristics and time of the discharge are shown in Figure 8.

Figure 9 (Plate II) shows the result with the same exposure conditions, with the condenser charged to 108 volts. The peak value of rate of growth of current is now 2.4×10^8 amp/sec. The actual values of current and rate of growth of current can be obtained from Figure 8 by multiplying by a factor 2.4. The time of the discharge is still as shown in Figure 8. In Figure 9 two spots have formed, disintegrating into semicircular lines moving with a velocity of nearly 10^4 cm/sec. Five positions are shown, the first appearing about 0.5 microsecond after the start of the discharge, so that just over 4 microseconds of discharge time are shown.

Figure 10(a) and (b) (Plate II) shows two photographs of spots obtained with the same exposure and circuit conditions as before, but with the condenser charged to 144 volts. The time of the discharge is still as shown in Figure 8, and actual values of current and rate of growth of current can be obtained by multiplying the values here shown by a factor of about 3.4. In (a) two spots have simultaneously formed and spread out into lines moving with an initial velocity of about 1.3×10^4 cm/sec. The cathode spot is shown for about the first five microseconds in about six positions—the first exposure of the right-hand spot occurring practically simultaneously with the formation of the spot, the left-hand spot having formed about 0.1–0.2 microsecond later.

In (b), three spots have formed simultaneously and spread, initially moving outwards at 10^4 cm/sec. The first few microseconds are shown in three or four exposures, the first exposure being about 0.1–0.2 microsecond after the start.

Hence we see that a spot can support a maximum rate of growth of current of about 10^8 amp/sec. (if it starts outwards from the wall of the tube). If higher rates of growth of current are required, then extra spots appear simultaneously. This is approximately what was predicted from the study of the spot in the magnetic field. If the spot should form away from the wall of the tube then the critical value would be 2×10^8 amp/sec.

The formation of fresh spots should not be confused with the appearance of large numbers of unit spots formed when the velocity of the emitting line decreases to slightly less than 10^4 cm/sec. These unit spots are not fresh spots, but broken-up pieces of emitting line. Fresh spots only form under the conditions of rapid growth of current. If the current is relatively slowly rising (as was the case in Figure 5) and unit spots have formed, the size of the emitting area increases by the breaking up of unit spots into greater numbers, and not by the spontaneous formation of fresh spots. If the current is relatively slowly rising, but the velocity of the emitting area is sufficiently high for unit spot formation not to take place (e.g. with the spot in a magnetic field), then the emitting area preserves its line-like structure, increasing proportionally with increasing current. This suggests that when unit spot formation has taken place each unit spot consists of a piece of line in the bottom of a dent, and that if the current is increasing the pieces of line in some of the dents increase, resulting in instability of the dents, so that they break into two or more separate dents. If the current is decreasing the total length of emitting line decreases proportionally, and the number of unit spots decreases.

Returning to the discussion of rapidly rising currents with the cathode spot not in a magnetic field, Figure 11 shows a plot of length of emitting line against current for a number of discharges similar to those shown in Figures 7 to 10

(Plate II). Again we find approximate linearity between the length of line, l , and current, i (amp.), such that $l = Ki$ cm., where $K = 3.7 \times 10^{-4}$ cm/amp $^{-1}$.

In view of the experimental difficulties, this value of K may be taken as identical with that obtained for the spot in a magnetic field.

§ 4. DISCUSSION AND CONCLUSIONS

It is a curious feature that, whereas for mercury (Froome 1949) the fresh spots form in front of the previous ones at high values of rate of growth of current, with the liquid alloy they almost invariably form along the edge of the cathode adjacent to the glass wall. Further, there is less time lag between the achievement of the high value of rate of growth of current and the formation of the fresh spots required to sustain it. For mercury this time lag may be of the order of a microsecond, whereas for the liquid alloy it appears to be nearer 0.1 microsecond.

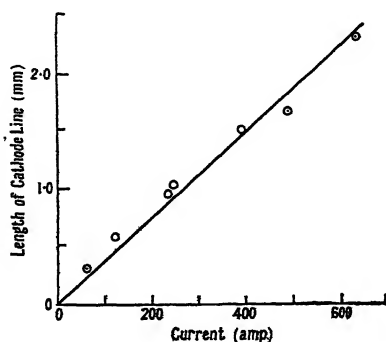


Figure 11.

Most of the conclusions given for mercury apply also to the liquid alloy; the most significant feature is that when the emitting area takes the form of a line, it is shorter than for mercury at the same current. The apparent width of the line for the alloy, as far as can be estimated, is between 10^{-3} and 5×10^{-4} cm., leading to an apparent emission current density of 2.5×10^6 to 5×10^6 amp/cm 2 , whereas the apparent current density for mercury is only slightly in excess of 10^6 amp/cm 2 .

There is considerably less copious evolution of vapour from the alloy cathode than from the mercury cathode, and it may be that for mercury the large amount of vapour emitted renders the surface at the cathode spot indeterminate and lowers its effective work function.

For the liquid alloy the emitting line is easily broken up into unit spots even when moving at a velocity of 10^4 cm/sec. For mercury the corresponding velocity is about 2×10^3 cm/sec. It is well known that the cathode spot exerts a considerable net downward pressure on the cathode under it, so that for the less dense liquid alloy rupture into dents (and hence unit spots) would take place more rapidly.

All the photographs here reproduced are of a region of dense and highly excited vapour shot off from the cathode at the emitting area. It is not known how far out from the cathode this region extends, or if there is a 'dark space' between this region and the cathode, as is the case for mercury (see Smith 1946). For mercury this 'dark space' extends outwards for about 10^{-3} cm., and as the

width of the emitting line appears also to be 10^{-3} cm. for mercury, the writer has suggested (Froome 1949) that the actual emitting line may be thinner than this and that a method of estimating its real size (as opposed to its apparent size) is to determine the maximum current which the spot will carry. For mercury the value is between 0.5 and 0.1 amp., and if at this value the spot is symmetrical, i.e. its length is equal to its breadth, the suggested real emission current density is 2×10^6 to 10^7 amp/cm². For the liquid alloy the minimum current lies between the same values; hence, by applying the same argument, the real emission current density may lie between the extraordinarily high values of 1.2×10^7 and 6×10^7 amp/cm², if we assume that for a current of i amp. the length of the emitting line l is given by $l = Ki$ where $K \simeq 4.0 \times 10^{-4}$ cm. amp⁻¹.

Even if we do not assume the real value of emission current density to be as high as that given above, and take only the apparent value of 2.5 to 5×10^6 amp/cm², we find that Langmuir's (1923) original theory of field emission is well substantiated without the assumption of an enhanced electric field at the cathode due to surface bumps. This is the first time that emission densities as high as those predicted by Langmuir's theory have been observed.

Langmuir suggested that since electrons have a high mobility, the positive ion current to the cathode, across the cathode fall of potential, must be positive space-charge-limited. This means that a certain field must exist in order that this current can pass, and he assumed that this field would be high enough to extract the electron current by 'cold' or 'high field' emission from the cathode surface.

Langmuir's well known space-charge formula reads

$$J_+ = \frac{1}{2\pi} \left(\frac{2e}{M_+} \right)^{1/2} \frac{V^{3/2}}{d^2}$$

where J_+ is the positive ion current density, V the cathode fall of potential, d the thickness of the cathode dark space, M_+ mass of the positive ions, and e the electronic charge.

To pass this current a field X is needed where

$$X = 4\pi \left(\frac{M_+}{2e} \right)^{1/4} J_+^{1/2} V^{1/4} = (3.28 \times 10^7)^{1/2} M_+^{1/4} V^{1/4} \text{ volt/cm.};$$

M is now the atomic weight of the positive ions, and J_+ and V are in practical units.

For the liquid alloy this reduces to $X = 2.6 \times 10^4 J_+^{1/2}$, assuming for convenience a cathode fall of 10 volts.

To obtain the electron emission density produced by this electric field, the equation of Fowler and Nordheim (1928) is best used. This states that the emission current density J_- for an electric field X is given by

$$J_- = \frac{6.2 \times 10^{-6} X^2}{\phi} \exp \left(\frac{-6.8 \times 10^7 \phi^{3/2}}{X} \right) \text{ amp/cm}^2$$

where ϕ , the work function, is 2 volts.

Figure 12 shows the electron emission density plotted against various incoming positive ion densities. It is seen that as soon as the positive ion current density attains a value of 10^6 amp/cm² an electron emission density greater than this value is obtained. The results described show apparent emission densities

of up to 5×10^6 amp/cm². At such a value a positive ion current density of 10^6 amp/cm² is quite conceivable, so that the results described may be taken to verify Langmuir's theory.

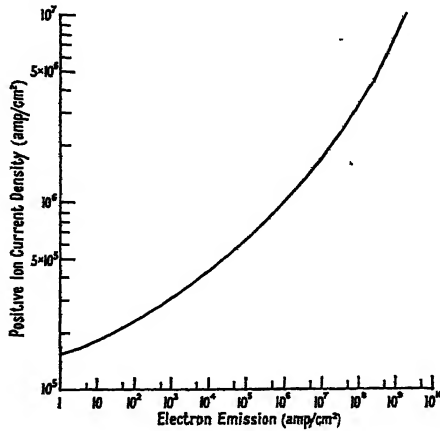


Figure 12.

Further, as for mercury, there is an apparent afterglow, lasting a microsecond or so, in the wake of a moving cathode spot.

ACKNOWLEDGMENTS

The writer wishes to express his thanks to Professor Sir George P. Thomson for many valuable discussions, and to Mrs. K. D. Froome for preparation of the diagrams.

REFERENCES

- FOWLER, R. H., and NORDHEIM, T., 1928, *Proc. Roy. Soc. A*, **118**, 229.
 FROOME, K. D., 1948, *J. Sci. Instrum.*, **25**, 371; 1949, *Proc. Phys. Soc. B*, **62**, 805.
 LANGMUIR, I., 1923, *Gen. Elect. Rev.*, **26**, 731.
 SMITH, C. G., 1946, *Phys. Rev.*, **69**, 96.
 THOMSON, J. J., and THOMSON, G. P., 1928, *Conduction of Electricity through Gases*, Vol. I (Cambridge: University Press), p. 479.

The Calculation of Magnetic Lens Fields by Relaxation Methods *

By M. B. HESSE†

Imperial College, University of London

*Communicated by G. I. Finch; MS. received 6th July 1949, and in amended form
26th October 1949*

ABSTRACT. The field and lens constants are calculated for two typical magnetic lenses as used in electron microscopes, using the relaxation method developed by Southwell for the solution of potential problems. The results are compared with those based on simple analytic approximations to the field distribution in the work of Glaser and Ramberg, and are shown to agree closely as regards those characteristics on which the discussion of the performance of the lens is based.

§ 1. INTRODUCTION

IT is an advantage in many problems connected with the design of electron lenses to have an accurate knowledge of the field distribution in the neighbourhood of the pole pieces. The determination of the characteristics of a lens, including the aberrations, can be carried out if the field distribution along the axis of symmetry of the lens is known. Bertram (1940, 1942) has solved the field equation analytically with boundary conditions approximating to those of a practical lens with unsaturated pole pieces, and finds a field expression of the form

$$H = H_m \operatorname{sech}^2 az.$$

Glaser (1941) has exhaustively studied the trajectories due to a field of the type

$$H = \frac{H_m}{1 + (z/a)^2}.$$

No attempt has, however, been made to solve theoretically the problem of the magnetic lens with saturated pole pieces, although this is the condition under which lenses are worked in practice.

This paper describes the solution of the field equations by the relaxation method developed by Southwell and his collaborators (Southwell 1946) for the numerical solution of potential problems. The calculation has been carried out for both unsaturated and saturated pole pieces.

§ 2. EQUATIONS FOR THE FLUX FUNCTION

In the first part of this paper it will be assumed that the relation between the magnetic field \mathbf{B} and the magnetizing force \mathbf{H} is linear, that is, $\mathbf{B} = \mu \mathbf{H}$, where μ is the constant magnetic permeability. This is only true over a restricted range of \mathbf{H} for materials used in practice, and the modifications required when μ is not constant will be discussed below.

The application of the relaxation method to a problem of field distribution is simplest when the field equations are expressed in a form containing a function ψ

* This work formed part of a thesis submitted to the University of London for the Ph.D. degree.

† Now at Royal Holloway College, University of London.

which is constant along the flux lines. ψ is defined by the equations

$$\mu H_z = -\frac{1}{r} \frac{\partial \psi}{\partial r}, \quad \mu H_r = \frac{1}{r} \frac{\partial \psi}{\partial z}. \quad \dots\dots (1)$$

The equation for ψ is

$$\frac{\partial^2 \psi}{\partial r^2} + \frac{\partial^2 \psi}{\partial z^2} - \frac{1}{r} \frac{\partial \psi}{\partial r} = \frac{0.4\pi nI}{\Delta} r, \quad \dots\dots (2)$$

where nI is the number of ampere turns in the coil and Δ is the total cross-sectional area of the windings.

Physically, the value of ψ at any point gives the magnetic flux passing through a circle which lies in a plane perpendicular to the axis, with its centre on the axis and its circumference passing through the given point.

At a boundary between media with different values of μ the tangential component of \mathbf{H} is continuous, and the normal component of \mathbf{B} is continuous.

Thus at a point on the boundary

$$(H_t)_1 = (H_t)_2; \quad (\mu H_n)_1 = (\mu H_n)_2. \quad \dots\dots (3)$$

From equations (1)

$$(\psi)_1 = (\psi)_2; \quad \left(\frac{1}{\mu} \frac{\partial \psi}{\partial n}\right)_1 = \left(\frac{1}{\mu} \frac{\partial \psi}{\partial n}\right)_2. \quad \dots\dots (4)$$

§ 3. THE RELAXATION METHOD

The solution of a potential distribution is given by the relaxation method in terms of a function defined by computed values, which satisfies the finite difference equation corresponding to the field at a finite number of points. The number may be increased indefinitely to obtain the degree of accuracy required, the only limitation being the labour involved in solving the system of simultaneous equations. The relaxation method is essentially a device for solving a large number of such equations quickly, and in such a way that mistakes can be detected and rectified at any stage.

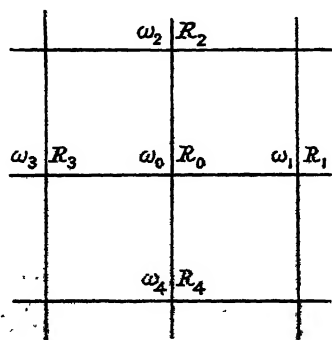


Figure 1.

The points at which the values of the function are to be calculated are most conveniently taken at the corners of a square mesh, as shown in Figure 1. The value of the function ω_0 at any general point of the field may be found in terms of the values $\omega_1, \omega_2, \omega_3, \omega_4$, at surrounding points, and the mesh length h . Special equations are needed at boundary points between media of different permeability.

We require to solve the differential equation for the flux function over the field of the lens. As the function will be computed in arbitrary units, equation (2) is replaced by

$$\frac{\partial^2 \omega}{\partial r^2} + \frac{\partial^2 \omega}{\partial z^2} - \frac{1}{r} \frac{\partial \omega}{\partial r} + rZ = 0. \quad \dots\dots (5)$$

ψ is then given by

$$\frac{\psi}{\omega} = \frac{0.4\pi nI/\Delta}{Z}, \quad \dots\dots (6)$$

where Z is a constant whose value depends on the units in which ω is calculated.

At ordinary points of the field the first-order finite difference equation corresponding to (5) is

$$\omega_1 + \left(1 - \frac{h}{2r}\right) \omega_2 + \omega_3 + \left(1 + \frac{h}{2r}\right) \omega_4 - 4\omega_0 + h^2 Z_0 r = 0.$$

This is obtained by expanding ω in a Taylor series at the point O and neglecting terms containing powers of h equal to or higher than the third. It is therefore accurate to the square of the mesh length.

The first step in the process of solution is to guess a distribution of ω throughout the field. Then the residual

$$R_0 = \omega_1 + \left(1 - \frac{h}{2r}\right) \omega_2 + \omega_3 + \left(1 + \frac{h}{2r}\right) \omega_4 - 4\omega_0 + h^2 Z_0 r \quad \dots\dots (7)$$

is calculated for each point of the mesh. The residuals are now reduced to zero by subtracting $R/4$ from ω at the point where R is greatest, and modifying the surrounding points accordingly. The process is repeated until all the residuals lie between -2 and $+2$ (where the units of this residual are those of the last digit or decimal of ω), after which no change in the values of ω will reduce them further. Values of ω which are fixed by the boundary conditions are of course never changed. Finally R is calculated from equation (7) for each point again, and any mistakes are detected by the occurrence of residuals greater than 2, and can be rectified by reducing these as before. At the final stage the signs of the residuals should be distributed at random over the field, and the sum of the residuals should be almost zero. This reduces the possibility of cumulative errors in any part of the field. The accuracy of the solution in any region can be increased by reducing the mesh length, or by taking account of higher order terms in the Taylor series when forming the difference equations.

Special difference equations are required for points near a boundary where the distances between neighbouring points are unequal, the so-called 'irregular stars'. Also it is convenient at some points to use a square net with sides inclined at an angle of 45° to the z axis. This is required near diagonal boundaries and in regions where a change-over to a net of smaller mesh takes place, and is also useful for interpolating the value at the centre of a coarse mesh. At boundary points it is assumed as a first approximation that the value of μ was everywhere of the order of 1,000, so that $1/\mu$ can be neglected in comparison with unity, giving a possible error of 0.1% in the value of ω . Then the boundary equation (4) reduces to $(\partial\omega/\partial n)_{\text{air}} = 0$. The region outside a sharp corner presents some difficulty. The flux lines tend to concentrate round the corner, and the value of the flux function varies rapidly so that the Taylor series converges slowly. If the corner were mathematically sharp the flux density would be infinite there.

but in fact the radius of curvature at the corner of the iron pole pieces is finite, and the mesh in this region should be at least as small as the radius.

The difference equations for all these special cases are given in the Appendix.

§4. CALCULATION OF THE LENS FIELDS AND TRAJECTORIES

The flux distribution was investigated for a simple lens such as might be used for the objective of an electron microscope, the permeability being assumed infinite. The boundary conditions of the problem are given by the results of the previous section, and the value of ω along the axis is given by

$$\frac{\partial \psi}{\partial r} = -\mu r H_z, \quad \frac{\partial \psi}{\partial z} = \mu r H_r$$

ψ and ω are therefore zero on the axis.

At the first stage of the calculation the value of Z was not defined, but a reasonable distribution of ω was assumed, and at each mesh point the function

$$\omega_1 + \left(1 - \frac{h}{2r}\right)\omega_2 + \omega_3 + \left(1 + \frac{h}{2r}\right)\omega_4 - 4\omega_0,$$

or the equivalent expression for boundary points, was calculated. Then the average value of this residual was obtained and Z chosen so that the total residual over the area of the windings was zero. This device was found to lighten the work considerably, since the only fixed values of ω at which an accumulated residual could be 'thrown away' lie on the axis at one corner of the field, opposite the air gap (marked X in Figure 2).

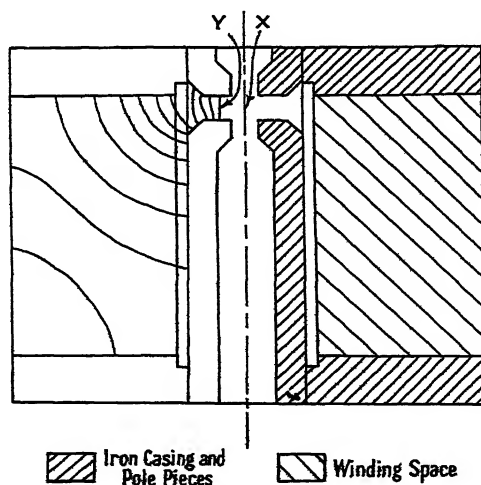


Figure 2. Section across objective lens I showing flux lines in the windings.

Rigorous justification of the method of solution is not necessary; the method is immaterial if the final values of ω satisfy the original set of linear equations, which have a unique solution. The error involved in approximating to the differential equation by finite difference equations, and the possible accumulation of errors in the relaxation process itself, however, has to be discussed. It was mentioned above that the finite difference equations are an accurate representation of the differential equation if powers of the mesh length higher than the third

can be neglected. This can be tested in the case of any solution by halving the mesh length and finding the values of ω at the new nodes. If the values previously obtained in the larger mesh remain unaltered, the neglect of terms in the Taylor series is justified; if not, the mesh length must be further reduced until two successive calculations correspond.

In unsaturated iron the permeability coefficient is about 1,000. Thus three-figure accuracy can be obtained in the values of ω if infinite permeability is assumed.¹ This accuracy can be maintained over the whole field irrespective of the absolute value of ω , as may be shown by considering two sets of solutions,

$$u_1, u_2, u_3, \dots u_n,$$

and

$$u_1 + \delta u_1, u_2 + \delta u_2, u_3 + \delta u_3, \dots u_n + \delta u_n,$$

where $\delta u_r/u_r < \epsilon$ and ϵ is the possible proportional error of u_r in one region of the field. The only fixed boundary values of ω are those on the axis, where ω is zero, and since the field represents a set of simultaneous linear equations in $\omega_1, \omega_2, \dots \omega_n$, any set of solutions may be multiplied in the same ratio to give another set of solutions. Thus

$$\frac{\delta u_1}{u_1} = \frac{\delta u_2}{u_2} = \dots = \frac{\delta u_n}{u_n},$$

and the possible percentage error remains constant over the whole field.

The absolute value of ω falls rapidly towards zero near the axis, so that the values of ω must be calculated to a greater number of decimal places near the axis than is necessary near the pole pieces (see Table 3), but since the possible percentage error remains constant, three-figure accuracy can be obtained throughout the field. Another possible source of error arises from the fact that the only fixed values of ω lie on the axis, and errors in the rest of the field are likely to be cumulative. The sum of the residuals in any arbitrary area of the distribution should be numerically less than 10% of the number of mesh points in order to reduce this possibility.

The absolute value of H_z along the axis may be found in terms of the number of ampere turns in the coil by an application of Ampère's law

$$\int_C \mathbf{H} \cdot d\mathbf{s} = 0.4\pi nI,$$

where C is a closed curve lying along the axis and round the outside of the iron case where the field is zero. Using equations (1) and expanding ω in a Taylor series we obtain

$$(h^2 H_z)_{r=0} = 2K\omega(h) \quad \dots \dots (8)$$

and

$$H_m \int_{-\infty}^{\infty} \frac{\omega(h)}{\omega_m} dz = 0.4\pi nI; \quad \psi = K\omega,$$

where ω_m is the value of $\omega(h)$ at $r=h, z=0$, and H_m is the value of H at $r=0, z=0$.

The distribution of H_z for lens I is shown in Figure 3.

The calculation for this lens indicated that one of the flux lines (Y in Figure 2) is straight and parallel to the axis at the centre of the pole-piece gap. If this is assumed to be generally true for all lenses with similarly shaped pole pieces within the limits of accuracy required, the work of calculating the field distribution

is considerably reduced. The distributions of H_z were found by this method for similar lenses in which the ratio of the gap to the inner diameter d of the pole pieces varies; these are shown in Figure 3, together with the distributions assumed by Glaser (1941) and Ramberg (1942).

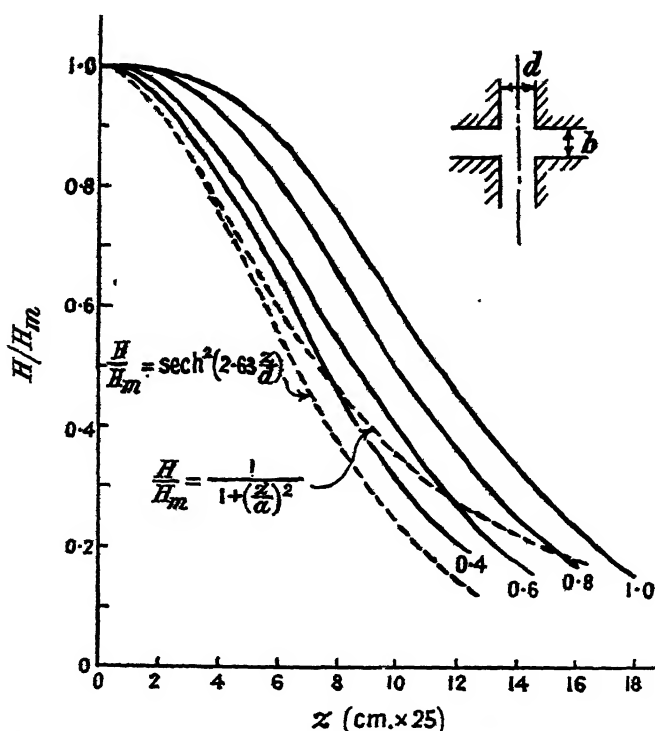


Figure 3. Variations of the field distribution with the pole piece ratio b/d .

The equation for the trajectories of paraxial electrons in a magnetic lens is

$$\frac{d^2 r}{dz^2} = \frac{-e}{8mc^2 V} r H^2. \quad \dots (9)$$

A numerical step-by-step method of solution is suitable, as the values of H are available at intervals along the axis. Neglecting terms of order h^4 , equation (9) becomes

$$r_{z-2h} + r_z - 2r_{z-h} = h^2 \frac{e}{8mc^2 V} r_{z-h} H_{z-h}^2.$$

Putting $H_{z-h} = H_m \frac{\omega(h)}{\omega_m}$ and $\frac{e H_m}{8mc^2 V} = k$,

we obtain
$$r_z = \left\{ 2 - k h^2 \frac{\omega^2(h)}{\omega_m^2} \right\} r_{z-h} - r_{z-2h}.$$

From this equation the radial distance of an electron is calculated at intervals: $z=h$. $\omega(h)/\omega_m$ is accurate to three figures, and five figures may be retained in the values of r_z . Since the number of steps is never greater than fifty, the final values of r_z may safely be regarded as accurate to three figures.

To find the focal length, the initial path of the electron before it enters the field is assumed to be at a constant distance from the axis. The focal length

is defined by analogy with geometrical optics. It is obtained by producing the final line of the path backwards to cut the initial path (Figure 4). s is the apparent shift of the centre of the lens system and f and s must be used in all calculations involving the formulae of geometrical optics. These formulae will only be valid if the object distance is greater than s .

Figure 5 shows the focal length f plotted as a function of $k = 6.85nI/\sqrt{v}$, representing the power of the lens.

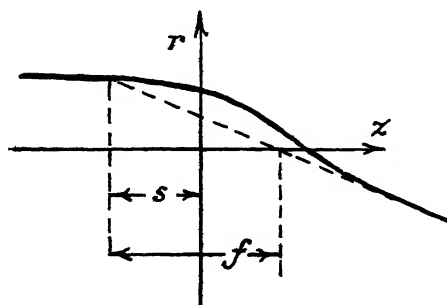


Figure 4.

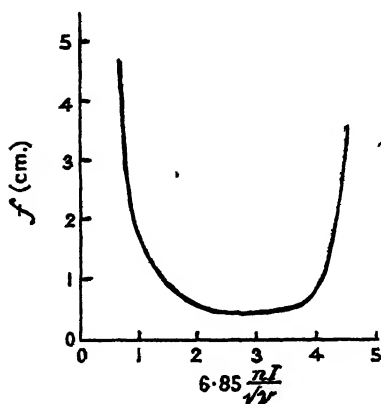


Figure 5. Variation of focal length with coil strength.

§ 5. CALCULATIONS FOR SATURATED LENS

The above results have been obtained by assuming that the permeability coefficient is greater than 1,000. This assumption is not valid for strong lenses, and its range of validity can be determined by finding the relation between the intensity of magnetization in any part of the iron and the number of ampere turns in the coil. In the lens under consideration it was found that saturation would occur if the number of ampere turns exceeded 1,200, and since the optimum strength of the lens has been found experimentally to lie between 4,000 and 5,000 ampere turns, it is evident that under normal working conditions the core and pole pieces are likely to be saturated. The possibility of using the relaxation method to calculate the field of a saturated lens was therefore examined.

If the iron in the field is saturated, the ratio B/H is not constant, and the relation $\mathbf{B} = \mu\mathbf{H}$ has to be replaced by

$$\mathbf{B} = f(\mathbf{H}),$$

where f is a function defined by the magnetization curve for the material. In an isotropic medium \mathbf{H} is parallel to \mathbf{B} , so that

$$H_r/B_r = H_z/B_z.$$

If ψ is defined in terms of \mathbf{B} as before, no simple equation can be found which is satisfied by ψ . It may, however, be assumed that the ratio B/H varies slowly over the field, and that in the region of any mesh point there is a linear relation between B and H . The coefficient μ_d is then defined as the slope of the tangent to the (B, H) curve. Thus $B - B_0 = \mu_d H$, where B_0 is the intercept of the tangent on the axis and μ_d and B_0 vary from point to point.

The function ψ may be defined as above by the equations

$$B_z = -\frac{1}{r} \frac{\partial \psi}{\partial r}, \quad B_r = \frac{1}{r} \frac{\partial \psi}{\partial z}.$$

Then equation (2) follows as before, and from conditions (3) the boundary equations are

$$\omega_1 = \omega_2 \quad \text{and} \quad \left(\frac{\partial \omega}{\partial n} \right)_1 = \mu \left(\frac{\partial \omega}{\partial n} \right)_2,$$

where $\mu = B/H$.

The process of solution is complicated by the fact that the values of μ are not known until B has been calculated, and μ is involved in the calculation of B . A method of successive approximation is therefore necessary. Bearing in mind the physical conditions of the problem, a distribution of μ along the boundary of the iron may be assumed and the flux function calculated as before, but using the exact expressions for the boundary conditions for finite μ . Only the equations for the boundary points are affected by the variation of μ ; those for the points inside the iron are exactly the same as for points in air (see Appendix). Given the flux function, B can be derived from equations (1), and hence new values of μ from the (B, μ) curve (Figure 6). After some experience of the behaviour of the distribution the computer can make the process converge fairly quickly.

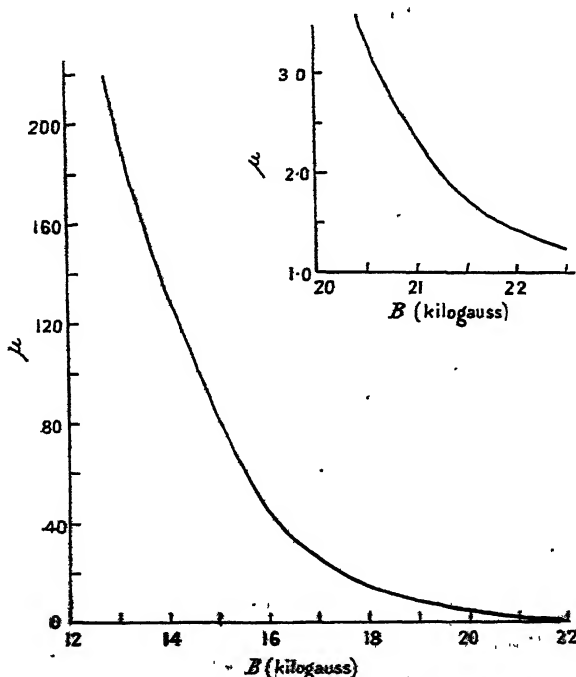


Figure 6. Variation of magnetic permeability with induction.

In order to investigate the field distribution when the pole pieces are saturated, the calculation was carried out for a lens similar to that illustrated by Hillier and Ramberg (1947, Figure 21) and used by them in attaining a resolution of 10 Å. The distribution for unsaturated iron was first calculated, the field being obtained

inside the iron as well as in air. Table 1 shows the distribution in the coil windings and Table 2 that in the region of the pole pieces on a larger scale. The field inside the iron in Table 2 has been obtained by assuming a uniform distribution of flux lines across the sections marked AA' and BB' in Table 1, these sections being taken at a sufficient distance from the pole pieces to avoid their disturbing effect on the uniformity. Table 3 shows the axial field distribution in more detail. The values of ω in these figures are given in arbitrary units; the units in Table 3 happen to be larger than those in Tables 1 and 2 by a factor 435/20. This calculation with infinite permeability showed that there is an area in the air gap where the flux lines are straight and parallel to the axis of the coil, that is, the line $\omega=20$ in Table 2. In such a region the radial boundary equation (see Appendix) with $\omega_1=\omega_3$ and $Z_0=0$ reduces to

$$\left(1 - \frac{h}{2r}\right)\omega_2 + \left(1 + \frac{h}{2r}\right)\omega_4 - 4\omega_0 = 0,$$

which is independent of μ . Therefore the introduction of a finite value of μ in this region does not affect the distribution, and it was assumed that saturation at the outer boundaries of the pole pieces could be regarded as having no effect on the region near the axis, and that the distribution of ω along lines surrounding the saturated tips was unchanged. Thus the problem was reduced to one with definite boundary values derived from the unsaturated lens. These boundary values are shown on the dotted line in Table 4. Nothing is assumed here about the absolute value of ω along this boundary; this is undoubtedly changed by saturation at the outer edge of the iron; it is merely the constancy of the distribution along the boundary that is required. The absolute value is determined by Z , which is involved in the conversion of ω into ψ (equation (6)). Since nothing is known about Z in this case, a different method has to be adopted in deriving the intensity from the function ω , which means that there is an uncertainty about the coil strength for which the calculation is being done until it is completed.

The process can best be explained by following the calculation step by step:

1. B_z, B_r are defined by the relations

$$B_z = \frac{K}{r} \frac{\partial \omega}{\partial r}, \quad B_r = -\frac{K}{r} \frac{\partial \omega}{\partial z}, \quad \dots \dots (10)$$

where $K\omega=\psi$, and K in this calculation is chosen arbitrarily. The order of magnitude of K is obtained by considering the corresponding constant in the unsaturated case and increasing it proportionately to the number of ampere turns for which the calculation is required.

2. A distribution of μ along the boundary is then guessed, and the function ω calculated over the field by the relaxation method, using the finite difference equations in the Appendix.

3. $\partial\omega/\partial z$ and $\partial\omega/\partial r$ can then be calculated for points on the boundary by differences, and hence B_z and B_r from equations (10). Then $B^2 = B_z^2 + B_r^2$, and a second approximation to μ at each point can be read off the (B, μ) curve.

4. Steps (2) and (3) are repeated until B and μ correspond within the limits of accuracy required.

The results for one coil strength are shown in Table 4.

Table 1. Field distribution in coil windings with infinite permeability.
The numerical values are those of ω , $\psi = 0.0162nI, \omega$

	511	510	506	500	492	482	471	460	450	442	437	435
509		508	504	498	490	480	468	456	446	437	432	430
504		503	499	493	484	473	460	447	435	425	418	415
498		497	492	485	475	463	448	433	418	405	396	393
491		490	485	477	466	452	435	417	398	381	368	364
485		483	478	470	457	441	422	401	377	353	341	334
480		478	472	463	450	433	412	387	356	338	321	305
									365	347	326	304
									351	337	324	314
									356	338	321	305
									365	347	326	304
									377	360	340	315
									376	359	338	311
									377	360	340	315
									376	359	338	311
									377	360	340	315
									376	359	338	311
									377	360	340	315
									376	359	338	311
									377	360	340	315
									376	359	338	311
									377	360	340	315
									376	359	338	311
									377	360	340	315
									376	359	338	311
									377	360	340	315
									376	359	338	311
									377	360	340	315
									376	359	338	311
									377	360	340	315
									376	359	338	311
									377	360	340	315
									376	359	338	311
									377	360	340	315
									376	359	338	311
									377	360	340	315
									376	359	338	311
									377	360	340	315
									376	359	338	311
									377	360	340	315
									376	359	338	311
									377	360	340	315
									376	359	338	311
									377	360	340	315
									376	359	338	311
									377	360	340	315
									376	359	338	311
									377	360	340	315
									376	359	338	311
									377	360	340	315
									376	359	338	311
									377	360	340	315
									376	359	338	311
									377	360	340	315
									376	359	338	311
									377	360	340	315
									376	359	338	311
									377	360	340	315
									376	359	338	311
									377	360	340	315
									376	359	338	311
									377	360	340	315
									376	359	338	311
									377	360	340	315
									376	359	338	311
									377	360	340	315
									376	359	338	311
									377	360	340	315
									376	359	338	311
									377	360	340	315
									376	359	338	311
									377	360	340	315
									376	359	338	311
									377	360	340	315
									376	359	338	311
									377	360	340	315
									376	359	338	311
									377	360	340	315
									376	359	338	311
									377	360	340	315
									376	359	338	311
									377	360	340	315
									376	359	338	311
									377	360	340	315
									376	359	338	311
									377	360	340	315
									376	359	338	311
									377	360	340	315
									376	359	338	311
									377	360	340	315
									376	359	338	311
									377	360	340	315
									376	359	338	311
									377	360	340	315
									376	359	338	311
									377	360	340	315
									376	359	338	311
									377	360	340	315
									376	359	338	311
									377	360	340	315
									376	359	338	311
									377	360	340	315
									376	359	338	311
									377	360	340	315
									376	359	338	311
									377	360	340	315
									376	359	338	311
									377	360	340	315
									376	359	338	311
									377	360	340	315
									376	359	338	311
									377	360	340	315
									376	359	338	311
									377	360	340	315
									376	359	338	311
									377	360	340	315
									376	359	338	311
									377	360	340	315
									376	359	338	311
									377	360	340	315
									376	359	338	311
									377	360	340	315
									376	359	338	311
									377	360	340	315
									376	359	338	311
									377	360	340	315
									376	359	338	311
									377	360	340	315
									376	359	338	311
									377	360	340	315
									376	359	338	311
									377	360	340	315
									376	359	338	311
									377	360	340	315
									376	359	338	311
									377	360	340	315
									376	359	338	311
									377	360	340	315
									376	359	338	311
									377	360	340	315
									376	359	338	311
									377	360	340	315
									376	359	338	311
									377	360	340	315
									376	359	338	311
									377	360	340	315
									376	359	338	311
									377	360	340	315
									376	359	338	311
									377	360	340	315
									376	359	338	311
									377	360	340	315
									376	359	338	311
									377	360	340	315
									376	359	338	311
									377	360	340	315
									376	359	338	311
									377	360	340	315
									376	359	338	311
									377	360	340	315
									376	359	338	311
									377			

The distribution of ω along the axis is then known, and since K has been arbitrarily chosen, the distribution of ψ is also known, and the value of nI for which the calculation has been done can be found from Ampère's law as follows:

From (8)
$$h^2 H_{r=0} = h^2 K \left(\frac{\partial^2 \omega}{\partial r^2} \right)_0 = 2K\omega(h).$$

Thus
$$\int H_z dz = \int \frac{2K}{h^2} \omega(h) dz = \frac{2K\omega_m A}{h},$$

where
$$A = \int \frac{\omega(h) dz}{\omega_m h}$$

and ω_m is the value of $\omega(h)$ at $z=0$.

But
$$\int H_z dz = 0.4\pi nI,$$

so that
$$nI = \frac{K\omega_m A}{0.2\pi h}.$$

When one calculation has been done for the highest value of nI required, the other distributions can be obtained easily by choosing the initial approximations

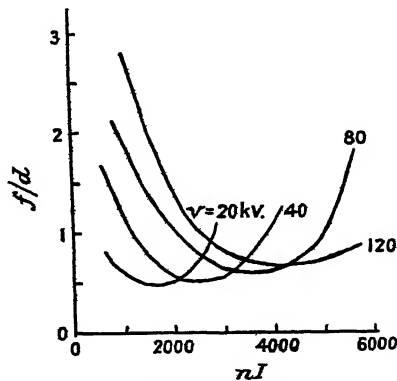


Figure 7. Variation of focal length with coil strength and voltage.

in the appropriate proportions. It was also sometimes useful to vary K in the middle of a calculation in order to reduce the residuals as rapidly as possible. The resulting coil strengths are shown in the table below.

K	1.12	1.68	1.95	2.18	2.35
ω_m	1.045	1.14	1.20	1.29	1.31
A	12.63	16.91	18.78	20.63	22.40
H_m (gauss)	5850	9570	11710	14090	15350
nI (amp. turns)	1180	2580	3560	4620	5480

When the coil strength was less than 1,000 ampere turns, it was found that the effects of saturation could be neglected, and, therefore, the distribution shown in Table 3 is valid for any coil strength below this limit.

The focal lengths of the lens for various coil strengths and accelerating voltages were calculated by the method of §4, and Figure 7 shows the curves of focal length plotted against ampere turns.



As before, the accuracy of the relaxation calculation to the number of figures required was ensured by halving the mesh length until the distribution remained unaltered by further increase in the number of nodal points. This also justified the use of discrete values of the variable μ at boundary points. The final accuracy of the results is limited by the knowledge available about the magnetic properties of the iron, but the value of μ was not found to be very critical, especially in regions where ω did not vary rapidly. A variation in μ of the order of 10% usually only produced a variation of 1% in ω . Thus if other sources of error are neglected, the error in ω is not greater than 1%, and it is possible to find the focal length to an accuracy of 0.1 mm.

§ 6. DISCUSSION OF RESULTS

Comparison has been made above with the 'classic' field distributions of Glaser and Ramberg. Cosslett (1946) has based a discussion of the resolving power of the electron microscope on their work, and it is of interest to compare this with the results obtained by the relaxation method.

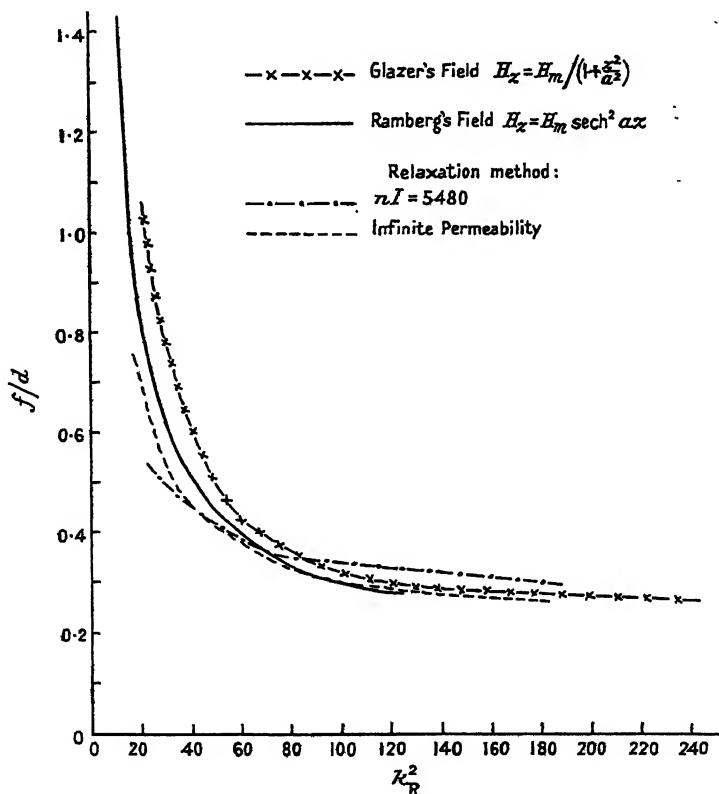


Figure 8. Variation of focal length with lens strength.

The curves of f/d against a parameter representing coil strength have been derived for both these distributions and are reproduced in Figure 8. The focal length f_R in this case has been calculated according to a different definition from that given above. The alternative definition is given by Ramberg in terms of

a trajectory initially parallel to the axis, and the slope of the trajectory at the point where it cuts the axis. Thus $f_R = r_0/\Delta_0$, where $\Delta_0 h$ is the slope at the point $r=0$ (Figure 4).

As can be seen from Figure 4, the two definitions are equivalent when the focal point is in field-free space. For strong fields f_R decreases asymptotically to zero, while the geometrical focal length becomes infinite, indicating the existence of two or more focal points for fields of this value. Ramberg's focal length is more convenient for the determination of the spherical aberration coefficient, although it assumes that the part of the field beyond the focal point does not contribute appreciably to the aberration. This assumption is justified in practice, since lenses are not worked at strengths beyond the minimum geometrical focal length, and in the working region the two focal lengths do not differ greatly.

In Figure 8 the abscissa k_R^2 is the lens parameter as defined by Ramberg, and is given by

$$k_R^2 = d^2 H_m^2 / (2.63)^2 V.$$

The Figure also shows the corresponding curves obtained by the relaxation method for the case of infinite permeability and for a coil strength of 5,480 ampere turns. The agreement is close, especially for values of k_R^2 between 60 and 100, and indicates that although the field distributions represented by Glaser's and Ramberg's expressions are not good approximations to the field of a practical lens, they do give a good idea of the variation of lens power with coil strength and accelerating voltage, and give a satisfactory basis for the calculation of the spherical aberration and the discussion of lens performance.

§7. CONCLUSIONS

During the fifteen years or so of the development of electron microscopes, progress in their construction has been made chiefly by trial-and-error methods, guided by mathematical treatment where approximations made this possible. A great many empirical generalizations have been made as a result of this work, and optimum operating conditions are now almost standardized. In view of the constant testing that these generalizations undergo in practice, it seems unlikely that the application of relaxation methods to the standard designs would lead to anything new at this stage. The relaxation method might however provide a useful means of investigating new designs, and one which avoids the practical difficulties of direct measurement.

APPENDIX

The finite difference equations to be satisfied at various types of points are as follows:

Ordinary points

$$\omega_1 + \left(1 - \frac{h}{2r}\right)\omega_2 + \omega_3 + \left(1 + \frac{h}{2r}\right)\omega_4 - 4\omega_0 + h^2 Z_0 r = 0.$$

Irregular stars

$$\frac{2\omega_1}{x_1(x_1+x_3)} + \frac{1}{x_2(x_2+x_4)} \left(2 - \frac{hx_4}{r}\right)\omega_2 + \frac{2\omega_3}{x_3(x_1+x_3)} + \frac{1}{x_4(x_2+x_4)} \left(2 + \frac{hx_2}{r}\right)\omega_4 - \left\{\frac{2}{x_1x_3} + \frac{2}{x_2x_4} + \frac{h}{r} \frac{x_2-x_4}{x_2x_4}\right\}\omega_0 + h^2 Z_0 r = 0.$$

Diagonal net—ordinary points

$$\left(1 - \frac{h}{2r}\right)(\omega_1 + \omega_2) + \left(1 + \frac{h}{2r}\right)(\omega_3 + \omega_4) - 4\omega_0 + 2h^2 Z_0 r = 0.$$

Diagonal net—irregular stars

$$\begin{aligned} & \frac{1}{x_1(x_1 + x_3)} \left(1 - \frac{hx_3}{2r}\right) \omega_1 + \frac{1}{x_2(x_2 + x_4)} \left(1 - \frac{hx_4}{2r}\right) \omega_2 + \frac{1}{x_3(x_1 + x_3)} \left(1 + \frac{hx_1}{2r}\right) \omega_3 \\ & + \frac{1}{x_4(x_2 + x_4)} \left(1 + \frac{hx_2}{2r}\right) \omega_4 \\ & - \left\{ \frac{1}{x_1 x_3} \left(1 + \frac{hx_1}{2r} - \frac{hx_3}{2r}\right) + \frac{1}{x_2 x_4} \left(1 + \frac{hx_2}{2r} - \frac{hx_4}{2r}\right) \right\} \omega_0 + h^2 Z_0 r = 0. \end{aligned}$$

Radial boundary points (Figure 9)

$$2\omega_1 + (\mu + 1) \left(1 - \frac{h}{2r}\right) \omega_2 + 2\mu \omega_3 + (\mu + 1) \left(1 + \frac{h}{2r}\right) \omega_4 - 4(\mu + 1) \omega_0 + h^2 Z_0 r = 0.$$

Cylindrical boundary points (Figure 10)

$$\begin{aligned} & \left(1 - \frac{h}{2r} + \frac{h\mu}{2r}\right) \omega_1 + 2 \left(1 - \frac{h}{2r}\right) \omega_2 + \left(1 - \frac{h}{2r} + \frac{h\mu}{2r}\right) \omega_3 + 2\mu \left(1 + \frac{h}{2r}\right) \omega_4 - \\ & 4 \left(1 - \frac{h}{2r} + \frac{h\mu}{2r}\right) \omega_0 + \left(1 + \frac{h}{2r}\right) h^2 Z_0 r = 0. \end{aligned}$$

Right-angled corner (Figure 11)

$$\omega_1 + \left(2\mu - 1 - \frac{h}{2r}\right) \omega_2 + (2\mu - 1) \omega_3 + \left(1 + \frac{h}{2r}\right) \omega_4 - 4\omega_0 + h^2 \mu Z_0 r = 0.$$

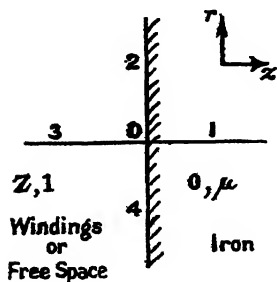


Figure 9.

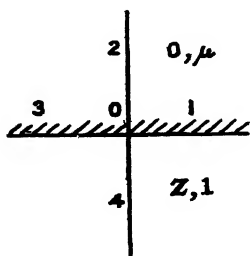


Figure 10.

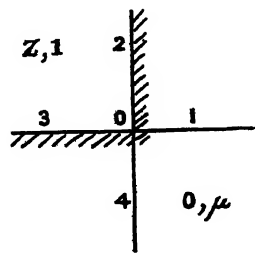


Figure 11.

ACKNOWLEDGMENTS

The author wishes to thank Professor G. I. Finch for permission to work in his laboratory and for his interest and advice during the course of the work, and also Mr. D. N. de G. Allen for assistance on several occasions.

REFERENCES

- BERTRAM, S., 1940, *Proc. Inst. Radio Engrs.*, N.Y., **28**, 418; 1942, *J. Appl. Phys.*, **13**, 496.
 COSSLETT, V. E., 1946, *Proc. Phys. Soc.*, **58**, 443.
 GLASER, W., 1941, *Z. Phys.*, **117**, 285.
 HILLIER, J., and RAMBERG, E. G., 1947, *J. Appl. Phys.*, **18**, 48.
 RAMBERG, E. G., 1942, *J. Appl. Phys.*, **13**, 582.
 SOUTHWELL, R. V., 1946, *Relaxation Methods in Theoretical Physics* (Oxford: University Press).

An Electrostatic Fluxmeter of Short Response-time for use in Studies of Transient Field-changes

By D. J. MALAN AND B. F. J. SCHONLAND

The Bernard Price Institute, University of the Witwatersrand, Johannesburg

MS. received 24th October 1949, and in final form 26th January 1950

ABSTRACT. The fluxmeter consists of a conducting system of small capacity which is alternately exposed to and screened from the electric field by the movement of a rapidly rotating metal shield, thus producing an alternating E.M.F. of period 0.83 millisecond. The output, after amplification, is displayed on the screen of a cathode-ray oscillograph and the record carries an automatic indication of the sense of the field every 7.5 milliseconds.

At the maximum sensitivity employed the device gives a deflection of one centimetre in a field of 20 volts/metre, with a background noise-level of 0.15 cm. (3 volts/metre).

The response-time of the instrument is one half-cycle (0.42 millisecond).

The fluxmeter has been developed to give the rapid response and high sensitivity required for studies of the electric fields of thunderstorms in the intervals between the separate strokes of a lightning discharge.

§ 1. INTRODUCTION

THE electrostatic fluxmeter described in this paper has been developed to provide a convenient method of measuring the electric fields produced by thunderstorms. The principle of the device was first described by Ross Gunn (1932) and it is sometimes referred to as a 'field-mill'. It does not appear to have been developed before in a form which gives it either the sensitivity or the rapid time of response of the instrument to be described here. It has been used by us in studies of the electric fields of thunderstorms in the interval of the order of 0.03 second between separate component strokes, and it has an obvious application in providing continuous records of electric fields over long periods. With a time resolution of somewhat less than one millisecond, it is capable of resolving field changes which the beautiful capillary electrometer method of C. T. R. Wilson (1916) misses, and in our experience it is more suitable for field strength records of long duration than the combination of exposed conductor and cathode-ray oscillograph devised by Appleton, Watson-Watt and Herd (1926), particularly when studying nearby storms. The latter method calls for high insulation resistance of the exposed conductor in order to maintain a long time-constant and this is a matter of some difficulty in the presence of heavy rain and point-discharge.

§ 2. DESCRIPTION OF THE INSTRUMENT

The instrument consists essentially of a conducting system of small capacity which is alternately and regularly exposed to, and screened from, the electric field by the movement of a rapidly rotating earthed metal shield. The exposure-screening cycle is carried out 1,200 times per second.

The conditions under which the cycle is performed cause an amplifying system connected to the underside of the conductor to give rise to an alternating E.M.F. of approximately sinusoidal form, whose amplitude is proportional to the strength of the field to be measured. This E.M.F. is amplified and passed to the X plates of a cathode-ray tube and the waveform is recorded by photographing the screen

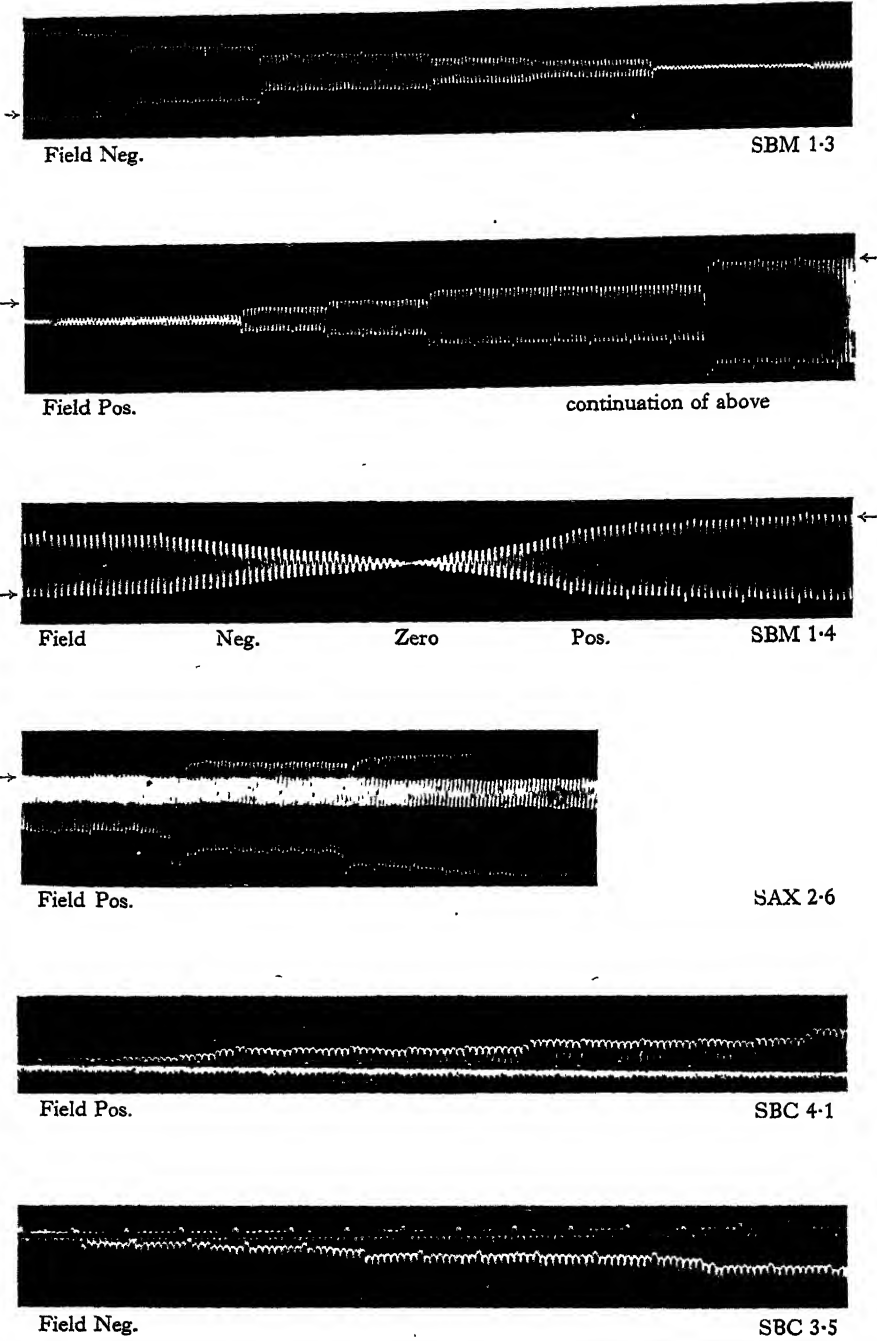


Figure 1.
Time scale: 1 cycle=0·8 millisc.

in a drum-film camera rotating about a horizontal axis. The amplitude of the envelope of the crests and troughs on this record is proportional to the field strength and alters as the field strength alters. A special device described in §5 indicates the sense of the field every ninth cycle and so shows which side of the envelope is to be selected for measurement. Some examples of records taken with the equipment using two different methods of recording are given in Figure 1 (see Plate). In all of these if the crests are enhanced in amplitude by 'polarity pips' the field is negative and vice versa. The arrows, therefore, show the side of the envelope which is used for measurement.

Constructional details of the instrument are shown diagrammatically in Figure 2, which is, however, not drawn to scale.

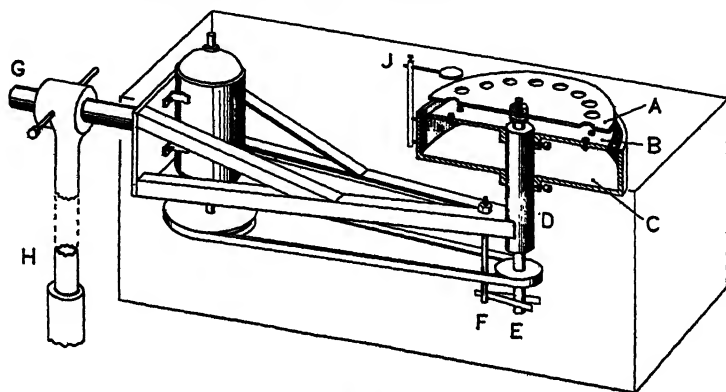


Figure 2.

The conducting system is formed by 18 metal studs mounted on a paxolin disc B. These studs, four of which are shown in the diagram, are constructed from No. 4 B.A. cheese-head screws of diameter 6.5 mm. and are set in a circle of diameter 20 cm. They are connected together below in the cylindrical brass box C and joined to the control grid of a low capacity triode valve V_1 (not shown) which is mounted inside C. The output of V_1 is led by cable to such further amplifiers as are necessary and from them to the cathode-ray tube. The capacity of the stud-grid system is 50 pF.

The screening disc A is of brass 23.5 cm. in diameter and 1.3 mm. thick. It is provided with 18 equally spaced holes of diameter 2.0 cm. whose centres lie immediately above the studs. The spacing between A and B can be adjusted by altering the position of B after loosening its clamping screws and is usually 2 mm. A is rotated at 4,000 r.p.m. by the steel shaft E, mounted on ball-bearings in the tubular housing D, and carrying a pulley which is belt-driven by a $\frac{1}{2}$ H.P., 1,400 r.p.m. A.C. motor.

The box C is mounted, like B, on the bearing housing D by means of a collar and clamping screws and by moving it downwards one can get at the valve without dismantling the two discs. In order to provide the best possible shielding for this valve, the box is provided with an earthed cover in the form of a thin metal sheet attached to the top of disc B and suitably perforated to give insulation to the studs.

The framework supporting the device is built of angle-iron with welded joints and carried on a horizontal pipe G, which enters a socket at the top of a vertical pipe H, one metre long, mounted on the roof parapet of the laboratory.

The whole framework is enclosed in a rectangular earthed metal case with the A side open and usually upwards. When it is necessary to run the instrument in rain or hail, it is inverted, so that A is downwards, and it is swung out over the outside wall of the laboratory so as to obtain a suitable ground clearance. Its sensitivity in the inverted position is reduced by a factor of two.

An essential part of the equipment is the earthing brush F, which consists of two springy phosphor-bronze strips pinching the shaft E between them. The surface of contact is slightly oiled and requires periodic cleaning: The purpose of F is to get rid of frictional charges which collect in an irregular manner on the disc A as a result of its rotation and of the belt drive. The brush and the whole framework are connected to earth. Unless this earthing is very good the noise-level of these frictional charges swamps the output.

To obtain a smooth output from the device, the discs A and B must be carefully constructed and quite flat. Any tendency to vibration in A must be eliminated since it gives rise to changes in capacity and so to an unwanted fluctuation in the output.

§ 3. AMPLIFYING CIRCUITS

Figure 3 shows the arrangement used in the head-amplifying circuit of V_1 . In order to obtain a low grid capacity, a Mullard E.F.37 pentode is connected up as a triode. This valve was chosen for its low microphonic properties, and as an extra precaution it is suspended by rubber bands. The effect of the grid resistor R_1 upon sensitivity is discussed in § 7. The output from V_1 is taken from a cathode-follower circuit and is passed to the recording room through a screened cable 7 metres long. The H.T. and filament supplies for V_1 are also provided through screened cable from the recording room and the whole equipment can thus be operated from a distance.

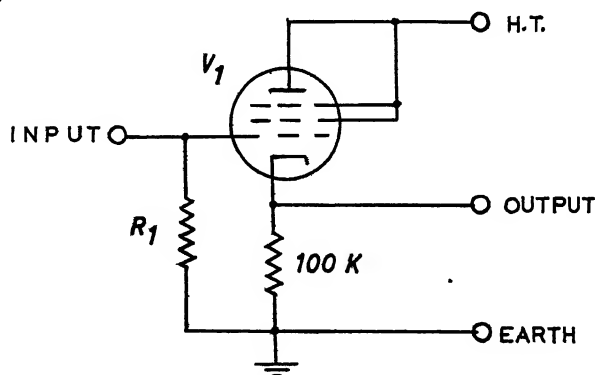


Figure 3.

Further amplification as required is provided by the X plate amplifier of the Dumont 208B oscillograph used to display the final output.

§ 4. METHODS OF RECORDING

In the present work we have been concerned with the recording of field changes associated with lightning flashes and have been able to use special trigger methods and fairly high film speeds. The drum camera used by us has a peripheral velocity of 44 cm/sec.

Two systems of recording have been employed. In the first, the time-base of the oscillograph is not used and the output is seen on the tube face as a horizontal line, which is screened off by a thin strip of black paper. An anticipatory lightning trigger (Schonland and Elder 1941) is arranged to operate in the initial stages of the stepped leader process of a flash to ground and to put a bias on the Y plates of the oscillograph before the leader has travelled very far. This bias causes the spot to move out from behind the screening strip and is kept on for one second so that the field changes taking place during the flash can be recorded. The method has the disadvantage that information about early first-leader field changes is lost, though it can be provided, if required, by another oscillograph kept continuously running. The first four records of Figure 1 are examples of this type of recording.

In the second method no trigger is used but the adjustable time-base of the oscillograph is employed. This time-base is 'locked' to the alternating E.M.F. from the studs so that the unmasked screen would show a single stationary wave pattern whose axis is vertical and whose amplitude alters with the field. A paper mask with a narrow horizontal slit is attached to the face of the screen in such a way as to cut out all but the crest of this stationary wave. When the field is steady or is varying slowly, the crest can be kept hidden behind a small masking disc at the centre of the slit, by manual adjustment of the X shift of the oscillograph. When the field alters rapidly, the crest travels along the slit and its displacement is recorded on the drum camera. In the case of many lightning flashes the sudden changes of field due to the leader process initiate the movement required for recording.

In Figure 1, SBC 4-1 and SBC 3-5 are examples of this type of recording; in the first the crest-spot has moved to a position above the masking disc, in the second it has moved below it.

§ 5. POLARITY INDICATOR

In order to show directly the sense or polarity of the electric field whose strength, as explained, is measured by the envelope of the amplitudes of the waveform, we have incorporated in the instrument a device which automatically increases the height of the crests of the waveform every ninth cycle if the field is negative in sense and the troughs every ninth cycle if it is positive. An inspection of the record immediately gives the required sense of the field.

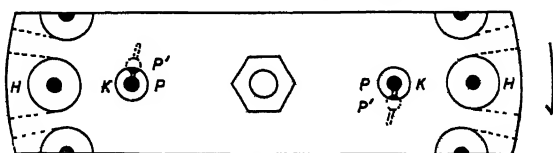


Figure 4.

Figure 4, which is a plan view of a portion of the disc A of Figure 2, shows how this is done. HH are the exposure holes in the screening disc, with the studs of disc B momentarily below their centres. An extra pair of holes, KK, have been placed in the positions shown. Below each of these are two extra studs P and P' fixed in slots in the paxolin disc B. Thin wires join these four 'polarity studs' to the output lead of the main studs. The phase of their contribution to the output can be adjusted by moving them in their slots.

When the disc A rotates in the direction of the arrow, the leading edges of the holes K and H uncover the peripheral studs and the studs P at the same moment, thus enhancing the output by the contributions from P. In the absence of studs P', both crests and troughs will be larger than before. The P' studs, however, become exposed when all other studs are being screened, so that their output is opposed in sense to that of the other studs. By adjustment of the position of the P' studs their output can be arranged to neutralize that of the P studs during the half-cycle that they are exposed. In this way the 'exposure' amplitude peaks of the output are larger than normal every half-turn of disc A but the 'screening' troughs are not affected. On reversal of the sense of the external field, the exposure half-cycle is that of the troughs and the polarity indication, as shown in the example marked SBM 1.4 (Figure 1) promptly shifts over to the other side of the zero line. All the records in Figure 1 show this polarity indication clearly. The easiest to follow is SBM 1.4 which begins with a negative electric field as shown by the enhanced crests and then passes through zero to an almost equal positive value, as shown by the enhanced troughs. This reversal of field occupies 0.06 second.

SBM 1.3, as shown by its enhanced crests at the start, begins with a large negative field which falls to nearly zero after five abrupt positive field changes caused by successive strokes of a lightning flash. In the record below the first the field is positive and five more strokes lead to a large positive field which ends the discharge. The sense of the fields in the other three examples shown in Figure 1 is similarly identified.

In practice it is necessary to adjust the height of the P studs with thin packing washers to bring them nearer to A and so to give them a slightly larger output than the P' studs.

§ 6. SENSITIVITY AND PERFORMANCE

The instrument in the form at present used by us gives a peak amplitude of 1 cm. in a field of 20 volt/metre and thus easily registers the electric field of fine weather which is of the order of 100 volt/metre. We have been able to record with it field changes due to separate lightning strokes at a distance of 25 km. The noise-level on our present site is about 1.5 mm. at the maximum amplification used and at the worst times of day. In a quieter locality the sensitivity could be much increased by using a larger disc with larger holes and larger studs or by cutting sectors from the disc, as shown by the dotted lines in Figure 4, and mounting the studs on a circle near to the edge of the disc. Local sources of electrical interference have, however, prevented us from using a higher sensitivity.

During rain when the arrangement is inverted and swung outwards, it gives half the above sensitivity when one metre from the building in a horizontal direction and 12 metres above the ground.

The equipment is calibrated by taking readings in fine weather and comparing them with simultaneous measurements by the stretched wire method in a nearby open field.

The electric fields of nearby thunderstorms are so large as to require reduction of the amplification provided in the oscillograph if the instrument is not to go 'off-scale'. This is easily met by adjustment of the calibrated attenuator in the oscillograph. It often happens, however, that there are charged clouds in the

neighbourhood when one wants to register small field changes due to distant storms and to use a high amplification. We have employed two methods to balance out such disturbing fields. The first makes use of a small insulated plate J (Figure 2), fixed to the box C and mounted above the rotating disc A. A controlled balancing potential can be applied to J to prevent the local field from driving the spot off-scale. The second method is to insulate the whole framework at the socket G of Figure 2 so that it and the disc A can be charged to an appropriate balancing potential while the studs and cathode follower are separately connected to earth as before.

Trouble was initially experienced with 50 c/s. hum from the A.C. mains but this was eliminated by balancing it against the output from a variable three-phase mains transformer.

§ 7. THE RESPONSE-TIME OF THE INSTRUMENT

Any sudden change ΔE in the electric field strength must produce a change $-\Delta Q$ in the charge on the upper surface of the system, given by

$$-\Delta Q = kA \Delta E / 4\pi,$$

where k is a site-correcting factor and A is the effective surface area of the conductor. A corresponding opposite charge ΔQ must appear on the under-surface. Provided the field change occurs during an exposure interval, the initial response of a device of this kind is therefore instantaneous. It can be shown, however, that the final steady value of the amplitude is less than the initial response and that a certain interval of time is needed for it to reach this steady amplitude after a sudden change.

Let us assume for simplicity that both exposure and screening intervals are equal to T and that both operations start and stop instantaneously. If C is the capacity of the system, including the grid of V_1 and the connecting leads, and R_1 is the grid resistor of Figure 3, we define α as e^{-T/CR_1} . It is then easily shown that successive half-cycles of exposure and screening give rise to changes in the charge on the under-surface of the conductor which are given by the terms of the converging series

$$\Delta Q, -\Delta Q(1-\alpha), \Delta Q(1-\alpha+\alpha^2), \dots \Delta Q[1-\alpha+\alpha^2 \dots + (-1)^n \alpha^n].$$

These terms converge to a final steady value $\Delta Q/(1+\alpha)$ when n is large, so that the final steady amplitude is $1/(1+\alpha)$ of that obtained from the first exposure 'kick', ΔQ .

In our usual arrangement, T is 0.40 millisecond, C is 50 pF. and R_1 is 2 megohms. α is then 0.0183 and the final steady amplitude is 1.8% less than the first 'kick'.

The response-time of the instrument will be defined as the number of half-cycles, N , before the amplitude reaches 99% of its final steady value. Since the n th term in the above series can be written

$$[1 + (-1)^n \alpha^{n+1}] \Delta Q / (1 + \alpha)$$

the number N is given by $\alpha^{N+1} \simeq 0.01$.

For the usual arrangement employed by us we have $\alpha^1 = 0.0183$, $\alpha^2 = 0.0033$ so that N should be one half-cycle.

Though the actual conditions of operation depart from the simple assumptions made in deriving this criterion, the instrument does, in fact, give a response-time

of one half-cycle with $R_1=2$ megohms. With a period of alternation of 0.83 millisecond, and a response-time of less than one complete cycle, the device therefore satisfies the requirements for which it was designed.

To obtain a higher sensitivity in observing distant thunderstorms we have occasionally increased R_1 to 10 megohms. Since in this case e^{-9T/CR_1} is 0.01, the response-time should be 8 half-cycles or 3.6 milliseconds. It is found in practice to be about 9 half-cycles.

With this value of R_1 , α becomes 0.449 and the final steady amplitude is only 0.69 of that given by the first half-cycle.

The amplitudes of successive half-cycles, as can be seen from the series given above, are not at first equal on the two sides of the zero. They only approach equality after the lapse of the N half-cycles of the response-time. Thus records taken with $R_1=10$ megohms, such as SAX2.6 of Figure 1, show an apparent shift of the zero when a sudden field change takes place. In practice this is sometimes useful in picking out small changes of field. The other records hardly show the effect, since they were taken with $R_1=2$ megohms and the zero shift after a field change lasts for 2 half-cycles at most.

ACKNOWLEDGMENTS

We wish to thank Dr. P. G. Gane for helpful advice on the development of this instrument, and Mr. V. Barnard and Mr. J. A. Keiller for their assistance in its construction.

REFERENCES

- APPLETON, E. V., WATSON-WATT, R. A., and HERD, J. F., 1926, *Proc. Roy. Soc. A*, **111**, 654.
 GUNN, R., 1932, *Phys. Rev.*, **40**, 307.
 SCHONLAND, B. F. J., and ELDER, J. S., 1941, *J. Franklin Inst.*, **231**, 39.
 WILSON, C. T. R., 1916, *Proc. Roy. Soc. A*, **92**, 555.

Wave Propagation in a Slipping Stream of Electrons : Small Amplitude Theory

By G. G. MACFARLANE AND H. G. HAY

Telecommunications Research Establishment, Ministry of Supply, Great Malvern, Worcs.

MS. received 8th December 1949

ABSTRACT. A theoretical study is presented of the TM-waves that can travel along a slipping stream of electrons. A slipping stream is defined as one in which the electrons move in parallel paths with velocity which varies with distance transverse to the motion. It is found that amplifying waves can travel along a slipping stream for all frequencies. It is also found that the slipping-stream tube has the remarkable property of combining the characteristics of a two-beam tube and a travelling-wave tube. This is due to the occurrence in the stream under suitable conditions of resonance layers, which act as highly reactive impedance sheets and can guide TM-waves of slow phase velocity in the same way as the spiral or corrugated surface in a travelling-wave tube.

The slipping-stream tube behaves differently according to the fractional velocity range of the electrons. If the velocity varies linearly from V_{-a} to V_a across the stream and $\alpha = (V_a - V_{-a}) / (V_a + V_{-a}) < 0.42$ it behaves primarily as a two-beam tube with plasma resonance frequency $\omega_0/2$, and therefore having a cut-off frequency $\omega_1 = \omega_0/\alpha\sqrt{2}$ and maximum gain $2.1\omega_0/V_0$ decibels per unit length, where ω_0 is the plasma resonance frequency of the slipping stream; in addition it has a low gain for $\omega > \omega_1$. However, if $\alpha > 0.42$ it behaves primarily like a travelling-wave tube and the rate of gain is about $0.53\omega_0/V_{-a}$ decibels per unit length at all frequencies above the plasma resonance frequency. This rate of gain is low compared with the gain possible with a travelling-wave tube in which the electron velocity is V_{-a} but it is achieved without the need for any external slow-wave waveguide.

Wave propagation along a slipping stream inside a waveguide, which in the absence of the electrons can guide a TM-wave of slow phase velocity, is also discussed. It is shown that this slipping stream travelling-wave tube has very much the same characteristics as a travelling-wave tube with a uniform electron beam. Maximum gain occurs when the phase velocity v_0 of the wave in the empty guide is about equal to the velocity of the electrons nearest to the reactive impedance sheet of the guide—in the example considered this is the fastest electron velocity V_a . A parametric set of curves is given relating the complex propagation constant to the frequency for different ratios v_0/V_a .

§ 1. INTRODUCTION

IN his original article on the electron-wave tube Haeff (1949) describes a tube in which a microwave signal can be amplified by interaction with a slipping stream of electrons. The slipping stream is a cylindrical beam in which the electrons are constrained to move in an axial direction, by the application of a strong steady axial magnetic field, with velocity which increases with radial distance from the axis of the beam. The stratification of velocity is obtained by surrounding the beam with a cylindrical conducting sheath or drift tube. The electron space charge within the drift tube depresses the potential along the axis, so that outer electrons travel faster than electrons along the axis (Smith and Hartman 1940). This results in what we shall call a slipping stream of electrons.

An analysis of the operation of this tube is not given by Haeff, although he does discuss a simplified theory of the two-beam tube in some detail. Experimental results are cited as proof that the slipping stream tube can amplify microwave signals. One of the most interesting characteristics of such a tube is the spectrum of frequencies for which amplification is possible, and this is not discussed by Haeff.

A steady slipping stream of electrons can also be maintained in an electron beam by the application of crossed electric and magnetic fields transverse to the axis of the beam. No strong axial focusing magnetic field is then required. Moreover when the electrons are perturbed by an oscillatory field they are no longer constrained to move in an axial direction but can oscillate transversely. However, to obtain parallel flow of the electrons in the non-oscillatory state the applied magnetic field must have a prescribed variation across the beam which depends on the electron density distribution, and this in turn is related to the velocity-distribution. If these values are not maintained the beam may alter in width along its length and the motion will cease to be single-stream. The non-oscillatory state of this system is the same as the single stream or type-S state of a linear magnetron in the particular example in which the electron density and the magnetic field are constant. In the more general case it would correspond to a type-S state of a magnetron with non-uniform electron density and magnetic field. In the oscillatory state, however, the periodicity of the field in the magnetron oscillator imposes the condition that the propagation constant must be real and the frequency, which is determined by the boundary conditions, must be complex for oscillations to build up in time, whereas in the slipping stream an oscillation of known real frequency impressed on the stream induces a wave to travel with increasing amplitude along the stream. The propagation constant, which is now determined by the boundary conditions, must therefore be complex. It is the slipping-stream tube with crossed electric and magnetic fields that we shall analyse in the following report.

Bunemann (1944) was the first to give a theory of the build-up of small amplitude oscillations from the type-S state in a magnetron oscillator and to derive an instability criterion. The analysis that we shall give for the slipping-stream amplifier will follow along much the same lines as that of Buneman; in particular we shall also use an action function to describe the electron flow.

One of the main objects of our work is to determine the spectrum of the frequencies which can be amplified. We shall consider not only a non-enclosed, or free, beam but also one which is loaded along its length by impedance sheets, which in the absence of the beam can guide a wave of slow phase velocity. The latter represents the beam inside a slow-wave waveguide in the manner of a travelling-wave tube. Although a loaded guide is not necessary for supporting an amplifying wave in a slipping stream of electrons it is of interest from the point of view of coupling the beam to the impressed field at the input and of extracting the amplified wave from the beam at the output. Also if ohmic losses in the walls of the guide are small it becomes apparent that a much higher rate of amplification is possible with a slipping-stream travelling-wave tube than with a free slipping stream.

§ 2. SMALL AMPLITUDE THEORY

(i) *An Arbitrary Velocity Distribution*

Our object is to investigate the properties of wave motions of small amplitude which can travel along a parallel beam of electrons, wherein the mean velocity varies with position across the stream. We use a quasi-stationary approximation throughout, in which retardation effects are neglected. For simplicity we consider only a two-dimensional variation in a cartesian coordinate system chosen so that there is no variation of any quantity in the direction of the x -axis. In the

non-oscillatory state the electrons are assumed to move parallel to the z -axis with a velocity which depends on y . The stream is therefore stratified parallel to the z -axis. We shall further assume that in the non-oscillatory state the electrons are constrained to move in parallel straight paths by the combined action of a magnetic field of vector potential $(0, A_y, 0)$ and of an electric field E_y . Then if \mathbf{V} is the electron velocity, ϕ the electric potential, ρ the charge density, we can obtain a non-oscillatory state for which $\text{curl } m\mathbf{V} = e\mathbf{B}$ by the use of the Hamilton-Jacobi equation (see Appendix I) and Poisson's equation. For this state,

$$\left. \begin{aligned} \mathbf{V} &= [0, 0, V(y)], \\ \mathbf{A} &= [0, A_y, 0], \\ \phi &= \frac{m}{2e} V^2(y) = \phi_0, \\ \rho &= \frac{\epsilon m}{2e} \left\{ V \frac{d^2 V}{dy^2} + \left(\frac{dV}{dy} \right)^2 \right\}. \end{aligned} \right\} \dots\dots(1)$$

Then an action function ψ_0 of the electrons exists, and is defined by

$$\text{grad } \psi_0 = m\mathbf{V} - e\mathbf{A}.$$

Moreover, the condition $\text{curl } m\mathbf{V} = e\mathbf{B}$ enables us to use the Hamilton-Jacobi equation, and is fulfilled if, as we assume in our case, the electrons leave the cathode with zero velocity and there is no component of magnetic field normal to the cathode surface (Appendix I).

Therefore

$$\left. \begin{aligned} \frac{\partial \psi_0}{\partial z} &= mV(y), \\ \frac{\partial \psi_0}{\partial y} &= -eA_y, \end{aligned} \right\} \dots\dots(2)$$

and the magnetic field is $[B_x, 0, 0]$, where

$$B_x = \frac{m}{e} \frac{dV}{dy}. \dots\dots(3)$$

The second equation of (2) defines the magnetic vector potential, from which the transverse magnetic field B_x is obtained.

It should be noted that both the charge-density distribution and the magnetic field, and hence the electric field E_y , are determined by the velocity $V(y)$.

We consider a perturbation in which $\text{curl}(m\mathbf{V} - e\mathbf{A})$ is still zero. Then we specify a possible case if we satisfy

$$\left. \begin{aligned} \text{grad } \psi &= m\mathbf{V} - e\mathbf{A} \\ e\phi &= \frac{1}{2m} (e\mathbf{A} + \text{grad } \psi)^2 + \frac{\partial \psi}{\partial t} \end{aligned} \right\} \dots\dots(4)$$

and

(which ensure that the equation of motion is satisfied), together with Poisson's equation and the equation of continuity. We consider a ψ of the form

$$\psi = \psi_0 + \psi_1(y)e^{i\theta},$$

where

$$\theta = \omega t - hz. \dots\dots(5)$$

In keeping with the restriction of the analysis to a quasi-stationary approximation, any perturbation of the magnetic vector potential will be neglected. Inserting (5) into (4) and retaining only first order perturbation terms we get

$$\phi = \phi_0 + \phi_1 e^{i\theta},$$

$$\text{where} \quad \phi_1 = \frac{i\omega}{e} \left(1 - \frac{hV}{\omega}\right) \psi_1. \quad \dots\dots (6)$$

$$\text{Poisson's equation is} \quad \frac{\partial^2 \phi}{\partial y^2} + \frac{\partial^2 \phi}{\partial z^2} = \frac{\rho}{\epsilon}. \quad \dots\dots (7)$$

We assume a small perturbation $re^{i\theta}$ on the charge density ρ . Then, from (7),

$$r = \frac{i\omega\epsilon}{e} \left[-h^2 \left(1 - \frac{hV}{\omega}\right) \psi_1 + \frac{d^2}{dy^2} \left\{ \left(1 - \frac{hV}{\omega}\right) \psi_1 \right\} \right]. \quad \dots\dots (8)$$

The next step is to eliminate r from (8) using the equation of continuity. Thus

$$\begin{aligned} J_z &= \rho V_z = (\rho + re^{i\theta}) \left(V - \frac{ih}{m} \psi_1 e^{i\theta} \right) \\ &\simeq \rho V + \left(Vr - \frac{ih}{m} \rho \psi_1 \right) e^{i\theta} \end{aligned} \quad \dots\dots (9)$$

$$\begin{aligned} \text{and} \quad J_y &= \rho V_y = (\rho + re^{i\theta}) \left(\frac{1}{m} \frac{\partial \psi_1}{\partial y} e^{i\theta} \right) \\ &\simeq \frac{\rho}{m} \frac{\partial \psi_1}{\partial y} e^{i\theta}, \end{aligned} \quad \dots\dots (10)$$

and the equation of continuity is

$$\frac{\partial J_z}{\partial z} + \frac{\partial J_y}{\partial y} = - \frac{\partial}{\partial t} (\rho + re^{i\theta}). \quad \dots\dots (11)$$

Therefore from (9), (10) and (11)

$$r \left(1 - \frac{hV}{\omega}\right) = \frac{i}{m\omega} \left\{ -h^2 \rho \psi_1 + \frac{d}{dy} \left(\rho \frac{d\psi_1}{dy} \right) \right\}. \quad \dots\dots (12)$$

On eliminating r from (8) and (12) we get the following differential equation for ψ_1 :

$$\begin{aligned} \frac{d}{dy} \left[\left\{ \left(\frac{\omega_0}{\omega} \right)^2 - \left(1 - \frac{hV}{\omega} \right)^2 \right\} \frac{d\psi_1}{dy} \right] + \left[\frac{h}{\omega} \left(1 - \frac{hV}{\omega} \right) \frac{d^2 V}{dy^2} \right. \\ \left. - h^2 \left\{ \left(\frac{\omega_0}{\omega} \right)^2 - \left(1 - \frac{hV}{\omega} \right)^2 \right\} \right] \psi_1 = 0, \end{aligned} \quad \dots\dots (13)$$

in which we have written

$$\omega_0^2 = \frac{e\rho}{\epsilon m} = V \frac{d^2 V}{dy^2} + \left(\frac{dV}{dy} \right)^2. \quad \dots\dots (14)$$

(ii) Linear Velocity Variation

We shall not make any attempt to find a general solution of (13) for ψ_1 but shall consider only a linear increase of velocity with y and put

$$V = V_0 + Qy. \quad \dots\dots (15)$$

Then from (1) $\rho = \epsilon m Q^2 / e$. Therefore from (14)

$$Q = \omega_0, \quad \dots\dots (16)$$

where ω_0 is the plasma resonance frequency.

On substituting

$$x = \frac{\omega}{\omega_0} \left(1 - \frac{hV}{\omega} \right) \quad \dots\dots(17)$$

into (13), the equation for ψ_1 reduces to

$$\frac{d^2\psi_1}{dx^2} + \frac{2x}{x^2-1} \frac{d\psi_1}{dx} - \psi_1 = 0. \quad \dots\dots(18)$$

Having found the equation that determines the motion of the electrons and the field inside the stream it is now necessary to ensure continuity of the field at the boundary of the stream. The next step is therefore to derive an expression for the radial admittance at the boundary of the perturbed stream. We use a method originally described by Hahn (1939).

If the boundaries of the unperturbed stream are $y = \pm a$ the boundaries of the perturbed stream oscillate about $y = \pm a$. For small perturbations it is sufficient to evaluate the magnetic field at the perturbed boundary $y = a + \Delta$ but to take this magnetic field as applying at the boundary $y = a$. In effect the perturbed boundary is replaced by the unperturbed boundary and the current between them is replaced by a surface current.

From Maxwell's equations we get an expression for H_x . Thus

$$\frac{\partial H_x}{\partial z} = \epsilon \frac{\partial E_y}{\partial t} - J_y. \quad \dots\dots(19)$$

Since $H_x = H_0 + H_1 e^{i\theta}$ \dots\dots(20)

and $E_y = -\partial\phi/\partial y$, it follows from (19) that

$$H_1 = \frac{\omega\epsilon}{h} \frac{\partial\phi_1}{\partial y} - \frac{i}{h} j_y, \quad \dots\dots(21)$$

where $J_y = j_y e^{i\theta}$.

If Δ is the vertical displacement of the boundary we have, neglecting the displacement current,

$$H_x(a + \Delta) - H_x(a) = J_z \Delta = \rho V_z \Delta. \quad \dots\dots(22)$$

Now, as shown in Appendix II,

$$\Delta = \frac{1}{m} \frac{\partial\psi_1}{\partial y} / i(\omega - Vh). \quad \dots\dots(23)$$

Hence on substituting from (21) and (23) into (22) we get

$$\begin{aligned} H_1(a + \Delta) &= \frac{\omega\epsilon}{h} \frac{\partial\phi_1}{\partial y} - \frac{i}{h} j_y - \frac{i\rho V}{m} \frac{\partial\psi_1}{\partial y} / (\omega - hV) \\ &= \frac{\omega\epsilon}{h} \frac{\partial\phi_1}{\partial y} - \frac{i\rho}{mh} \frac{1}{1 - (hV/\omega)} \frac{\partial\psi_1}{\partial y}. \end{aligned} \quad \dots\dots(24)$$

The second line of (24) follows from (10).

Also $E_z = -\partial\phi/\partial z = E_1 e^{i\theta}$. Therefore

$$E_1 = ih\phi_1 = -\frac{\omega h}{e} \left(1 - \frac{hV}{\omega} \right) \psi_1. \quad \dots\dots(25)$$

Finally, the wave admittance in the positive direction of y is given by

$$Y = \frac{H_1}{E_1} = -\frac{i\omega\epsilon}{h^2} \frac{d}{dy} \log \phi_1 + \frac{i\omega\epsilon}{h^2} \left(\frac{\omega_0}{\omega} \right)^2 \frac{1}{(1 - hV/\omega)^2} \frac{d}{dy} \log \psi_1. \quad \dots\dots(26)$$

If we normalize this admittance by substituting

$$P = -\frac{i h Y}{\epsilon \omega} \quad \dots\dots (27)$$

and change the variable to x using (17), (26) becomes

$$P(x) = \frac{1}{x} + \left(1 - \frac{1}{x^2}\right) \frac{1}{\psi_1} \frac{d\psi_1}{dx}. \quad \dots\dots (28)$$

We also require the value of the admittance of the field outside the stream at the stream boundaries. In the first instance we shall take the beam to be isolated in free space. The field outside the beam must then fall off in amplitude with distance from the beam. It is sufficient to represent this field by a simple evanescent TM-wave with phase velocity ω/h in the z direction. The admittance of such a wave is

$$Y = i \sqrt{\left(\frac{\epsilon}{\mu}\right) \frac{k}{\sqrt{(h^2 - k^2)}}}.$$

In the quasi-stationary approximation which we are using, $h \gg k$, so we take

$$Y = i \sqrt{\left(\frac{\epsilon}{\mu}\right) \frac{k}{h}}. \quad \dots\dots (29)$$

Normalizing it as in (27) gives

$$P(a) = 1. \quad \dots\dots (30)$$

Similarly the admittance at $y = -a$ is

$$P(-a) = -1. \quad \dots\dots (31)$$

(iii) *Determination of the Propagation Constant for the Free Stream*

The propagation constant h is determined by matching the admittance P at the boundaries of the stream.

Suppose we have found two independent solutions of the equation (18) for ψ_1 . Denote them by $F(x)$ and $G(x)$. Then the general expression for ψ_1 is

$$\psi_1(x) = c_1 \{F(x) + cG(x)\}. \quad \dots\dots (32)$$

From (27) the admittance is

$$P(x) = \frac{1}{x} + \left(1 - \frac{1}{x^2}\right) \frac{F'(x) + cG'(x)}{F(x) + cG(x)} \quad \dots\dots (33)$$

and the boundary conditions are

$$P(x_a) = -P(x_{-a}) = 1, \quad \dots\dots (34)$$

where, by (15) and (17),

$$\left. \begin{aligned} x_a &= \frac{\omega}{\omega_0} - \frac{hV_0}{\omega_0} \left(1 + \frac{a\omega_0}{V_0}\right) \\ x_{-a} &= \frac{\omega}{\omega_0} - \frac{hV_0}{\omega_0} \left(1 - \frac{a\omega_0}{V_0}\right) \end{aligned} \right\} \quad \dots\dots (35)$$

It will be seen from (35) that the unknown propagation constant and the frequency both appear in the expressions for the boundaries $x_{\pm a}$. This complicates the procedure for evaluating h .

Before we discuss the results of solving (33) and (34) we shall complete the analysis by considering a stream loaded with admittance sheets at the boundaries.

(iv) *Linear Velocity Variation with Loaded Guide*

In order to study the interplay of the electron space charge and a slow TM-waveguide wave we shall analyse the following system. The mean velocity of electrons increases from 0 at $y = -a$ to $2\omega_0 a$ at $y = +a$. At the boundary $y = -a$ we assume a perfectly conducting sheet and at $y = a$ an impedance sheet with such properties as are necessary for the propagation of a wave of phase velocity v_0 between $y = -a$ and $y = +a$ in the absence of free electrons.

For this electron-free wave

$$\left. \begin{aligned} E_z &= E_0 \sinh T(y+a), \\ H_x &= -\frac{i\omega\epsilon}{T} E_0 \cosh T(y+a), \\ Y(a) &= -\frac{i\omega\epsilon}{T} \coth(2aT), \end{aligned} \right\} \dots\dots(36)$$

where $T^2 = h_0^2 - k^2$ and $h_0 = \omega/v_0$.

For $v_0 \ll c$
$$Y(a) = -\frac{i\omega\epsilon}{h_0} \coth(2ah_0). \dots\dots(37)$$

The normalized admittance at $y = a$ is therefore

$$P(x_a) = -\frac{h}{h_0} \coth(2ah_0). \dots\dots(38)$$

Consequently from (37), taking the velocity of electrons at $y = -a$ to be zero so that $V_0 = \omega_0 a$,

$$P(x_{-a}) = \infty$$

and

$$P(x_a) = -\frac{h}{h_0} \coth\left(\frac{2h_0 V_0}{\omega_0}\right).$$

Let η be the ratio of the phase velocity of the wave in the empty guide to the mean velocity of the fastest electrons. Then

$$\eta = \frac{v_0}{2V_0} = x_{-a} \frac{\omega_0}{2h_0 V_0}, \dots\dots(39)$$

and from (37)
$$h = \eta h_0 \left(1 - \frac{x_a}{x_{-a}}\right), \dots\dots(40)$$

where

$$x_{-a} = \frac{\omega}{\omega_0}.$$

The boundary conditions then become

$$\left. \begin{aligned} P(x_{-a}) &= \infty, \\ P(x_a) &= -\eta \left(1 - \frac{x_a}{x_{-a}}\right) \coth\left(\frac{x_{-a}}{\eta}\right), \end{aligned} \right\} \dots\dots(41)$$

where

$$P(x) = \frac{1}{x} + \left(1 - \frac{1}{x^2}\right) \frac{F'(x) + cG'(x)}{F(x) + cG(x)}.$$

From the first condition of (41) we deduce that

$$c = -\frac{F(x_{-a})}{G(x_{-a})}. \dots\dots(42)$$

§ 3. THE RESULTS

The analysis presented in § 2 is based on a quasi-stationary approximation of which the main assumptions are (a) that the phase velocity of the propagated wave is much less than c , (b) that the effect on the electrons of their own magnetic field is negligible in comparison with the effect of the externally applied magnetic field, (c) that the oscillating field is of small amplitude compared with the D.C. field. In the results to be described in this section attention has been focused on only one mode of wave propagation. It is the lowest-order mode, which in most applications would be the dominant mode. There is, however, an infinite set of higher-order modes, some of which may travel against the stream and some may be amplifying. We shall not pursue their study further in this work.

(i) *Solutions of the Equation for the Action Function*

The key to the behaviour of the slipping stream as an amplifier is contained in the functions $F(x)$ and $G(x)$, two independent solutions of the wave equation for ψ_1 , equation (18). Since x depends on h , which is in general complex, we must study F and G for complex values of the argument. We have therefore tabulated $F(x)$ and $G(x)$ defined in the following way:

$$\begin{aligned} F(1, 0) &= 1, & F'(1, 0) &= 0, \\ G(x) &= c_2[F(x) \log c_1(1-x) + g(x)], & g(1, 0) &= 0, \\ G(x) &= F(-x), & G'(x) &= -F'(-x). \end{aligned}$$

Then $F(x) = c_2[G(x) \log c_1(1+x) + g(-x)]$. The property of these functions which is of primary interest to us in this discussion is that $F(x)$ and $G(x)$ have logarithmic singularities at the points $(-1, 0)$ and $(1, 0)$ respectively. If we refer to the definition of x at equation (17) we find that at the singularities,

$$\omega = \frac{\pm \omega_0}{1 - hV/\omega}.$$

But the right-hand side of this equation is the plasma frequency ω_0 increased by the Doppler factor due to its motion with velocity V in a medium in which waves travel with velocity ω/h . The singularity therefore corresponds to a condition of resonance between the applied frequency ω and the apparent plasma frequency measured by a stationary observer.

(ii) *Properties of the Lowest-order Modes in a Free Slipping Stream*

In the analysis and discussion of the results it is convenient to describe the velocity range of the electrons in terms of $\alpha = (V_a - V_{-a})/2V_0$. When $\alpha = 1$ the velocity is zero at one edge of the beam and $2V_0$ at the other edge.

In Figure 1 we show the imaginary part of the propagation constant h as a function of frequency for three values of α . These results were obtained by solving the equations (33) and (34) of § 2(iii). The frequency axis is marked in units of the plasma resonance frequency ω_0 and h is measured in units of ω_0/V_0 .

Consider firstly the spectrum for $\alpha = 0.25$. Figure 1 shows apparently three amplifying and three attenuating waves, taking the minus and plus signs of $\text{Im}(hV_0/\omega_0)$ respectively. (In the following discussion we shall confine our attention to the amplifying waves only.) The wave marked (1) has an amplification range $0 < \omega < \omega_1$, and a dependence of gain on frequency which bears a

strong resemblance to that of a two-beam tube. Comparison of the two shows that the gain-frequency characteristic of the slipping-stream tube is very nearly the same as that of a two-beam tube which has stream velocities V_a and V_{-a} but plasma resonance frequency $\omega_0/2$. Since the Larmor frequency ω_L is equal to $\omega_0/2$, it appears that the plasma resonance frequency of the two-beam tube, whose electron space-charge field is balanced by an electric field of positive charges, should be compared with the Larmor frequency of the slipping-stream tube, whose space-charge field is balanced by a transverse magnetic field. As in the two-beam tube the phase velocity is about equal to the mean velocity V_0 of the

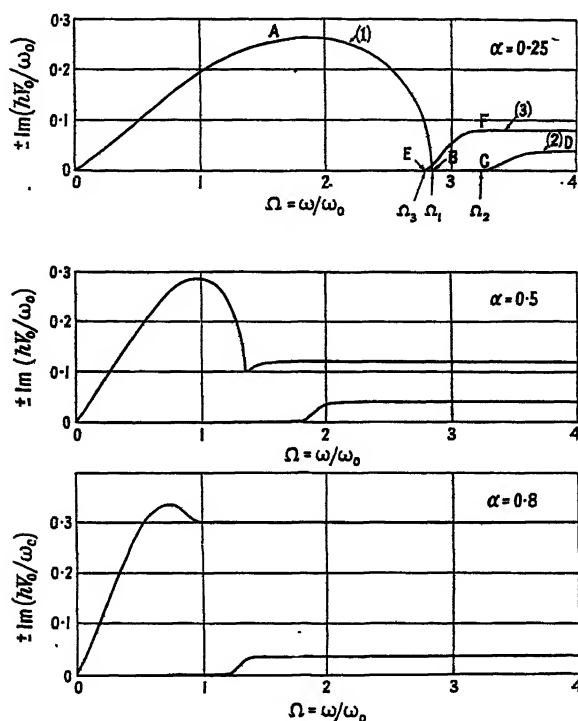


Figure 1. Dependence of rate of gain on frequency in a free slipping stream.

electrons. Thus wave (1) amplifies by the same mechanism as the dominant amplifying wave of the two-beam tube. It is therefore convenient to call it the two-beam tube wave.

Waves (2) and (3), however, have no counterpart in the two-beam tube. On the contrary they are of the same type as the amplifying wave in a travelling-wave tube. The effective reactive impedance sheet, which guides the slow wave, is provided by a resonance layer within the stream itself. In order to appreciate this phenomenon the reader is referred to Figures 2 to 5.

In Figure 2 the loci of the points x_a and x_{-a} , defined by (35), are plotted for a range of values of α . They are values of the variable x , which varies linearly across the beam from x_{-a} at the low-velocity edge to x_a at the high-velocity edge. In order to assist in the explanation by the following argument a sketch of the loci of x_a and x_{-a} for $\alpha \approx 0.25$ is given in Figure 3.

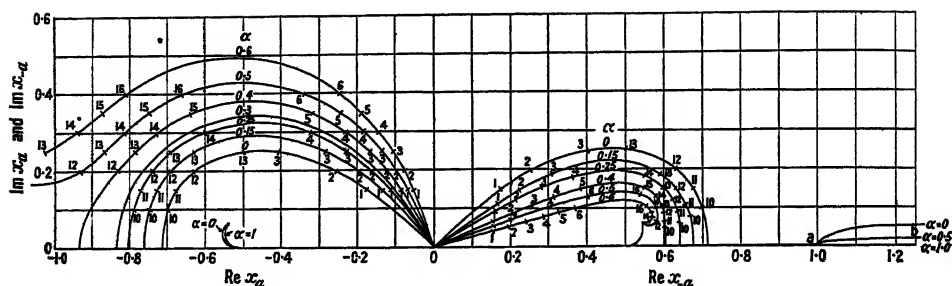


Figure 2. The loci of the points x_α and $x_{-\alpha}$ for a range of values of $\alpha = (V_\alpha - V_{-\alpha})/2V_0$. The numbers mark corresponding points of x_α and $x_{-\alpha}$.

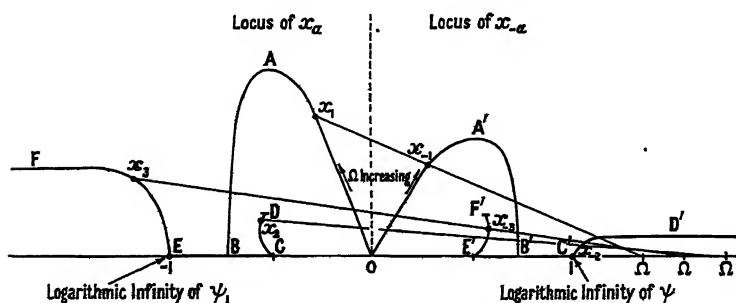


Figure 3. Sketch of the loci of x_α and $x_{-\alpha}$ for $\alpha < 0.42$.

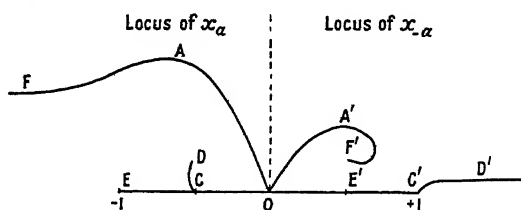


Figure 4. Sketch of the loci of x_α and $x_{-\alpha}$ for $0.42 < \alpha < 0.9$.

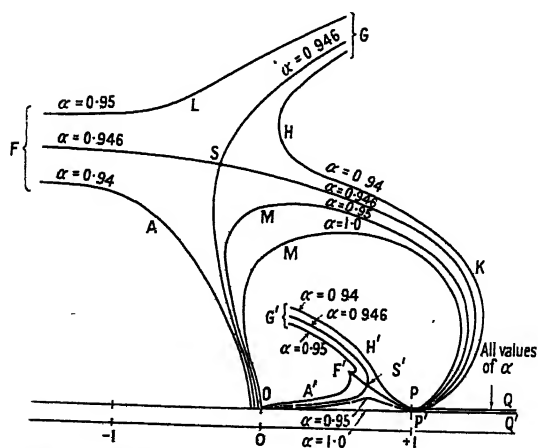


Figure 5. Sketch of the loci of x_α and $x_{-\alpha}$ for $0.9 < \alpha < 1.0$.

From the expression for x , namely

$$x = \frac{\omega}{\omega_0} - \frac{h}{\omega_0}(V_0 + \omega_0 y),$$

it follows that the value of x at the point of intersection of the straight line joining x_a to x_{-a} and the real axis is the frequency $\Omega = \omega/\omega_0$ and that

$$\text{Im}(hV_0/\omega_0) = \frac{1}{2}(\text{Im } x_a + \text{Im } x_{-a}).$$

The straight line joining x_a and x_{-a} also defines the distribution of the action function ψ_1 across the beam. Now, as indicated in §3(i) ψ_1 has logarithmic singularities at the points $(-1, 0)$ and $(1, 0)$. If, therefore, the line passes near to either of these points ψ_1 , and consequently E_{1z} , will have a pronounced maximum at the level in the stream that corresponds to the nearest point of approach. Moreover it is known (Ramo 1939) from the study of wave propagation in a uniform electron stream of velocity V that a slow TM-wave can propagate along the stream provided $\text{Re}(\omega - Vh) \simeq \pm \omega_0$. In terms of x this condition is $\text{Re } x \simeq \pm 1$. It signifies the state of resonance between the applied frequency and the plasma resonance frequency as measured by a stationary observer. If the real parts of x_{-a} and x_a enclose the point $x=1$, or the point $x=-1$, the resonance condition for the propagation of a slow TM-wave will be approximately satisfied at one level in the stream, which can be appropriately called the resonance level. The resonance level behaves therefore like the corrugated surface or reactive impedance sheet of a travelling-wave tube: it can guide a slow wave. The axial electric field falls off rapidly away from the resonance level in just the same way as from the effective impedance sheet of a travelling-wave tube. If therefore a resonance layer occurs in the beam of a slipping-stream tube we should expect the wave supported by it to interact with electrons in other levels in the beam in much the same way as the wave guided by the helix or corrugated structure reacts with the electrons in a travelling-wave tube.

In the light of these remarks let us consider how the loci of x_a and x_{-a} move as the frequency is increased from zero. In Figure 3 x_a and x_{-a} lie at the origin when $\Omega = \omega/\omega_0 = 1$. As Ω is increased x_a moves out along the curve OAB and x_{-a} along OA'B'. For these waves the real parts of x_a and x_{-a} do not enclose either $x=1$ or $x=-1$ and so no resonance layer occurs in the beam. This gives the two-beam tube wave of Figure 1. However, when x_a reaches the point B its path forks. One branch moves along the real axis to C then up to the limiting point D, where $x = -0.542 + 0.061i$, while x_{-a} moves to the resonance point C' and then out along C'D'. In the range BC the propagation constant is entirely real but it acquires an imaginary part beyond C, as shown for $\alpha = 0.25$ in Figure 1. The cut-off frequency at C is simply related to α by the formula

$$\Omega_2 = (0.754/\alpha) + 0.247.$$

In the range C to D it is clear from Figure 3 that a resonance layer occurs in the stream. It appears at the low-velocity edge of the stream at frequency Ω_2 and moves towards the high-velocity edge as the frequency is increased. The phase velocity also increases, as shown by the curve CD in Figure 6, in keeping with the velocity of the electrons at the resonance layer. As Ω tends to infinity $\text{Im}(hV_0/\omega_0)$ tends to $0.061/(1+\alpha)$, which gives a gain of $0.53\omega_0/V_a$ decibels per unit length.

If we now return to the point B and follow the other branch we find that x_a moves along the real axis to the singularity at E and then along EF, while x_{-a} moves to E', where $x = 0.507$, and up to F', where $x = 0.542 + 0.061i$. Figure 1 shows that as x_a moves from B to E the frequency decreases to $\Omega_3 = (0.754/\alpha) - 0.247$ and $\text{Im } h = 0$, but as x_a moves out along EF the frequency again increases and $\text{Im}(hV_0/\omega_0)$ increases asymptotically to the value $0.061/(1-\alpha)$ as $\Omega \rightarrow \infty$. This gives a gain of $0.53\omega_0/V_{-a}$ decibels per unit length. It is evident from Figure 3 that in the range EF a resonance layer occurs in the beam and that it moves from the high-velocity edge towards the low-velocity edge as the frequency increases from Ω_3 . The phase velocity also decreases, as shown by the curve EF in Figure 6.

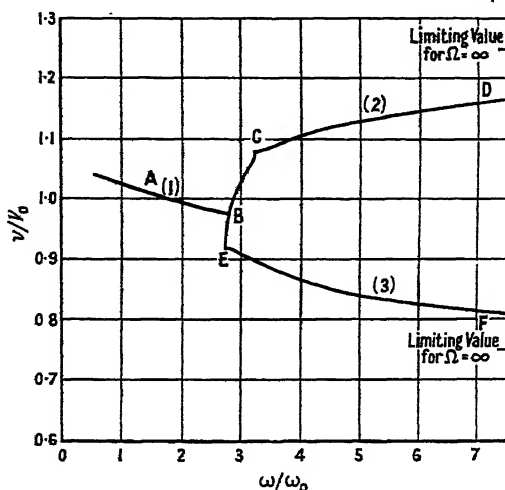


Figure 6. The phase velocity as a function of frequency for $\alpha = 0.25$.

We can now explain why wave (3) has a higher gain than wave (2) for $\Omega \gg \Omega_2$. It is because the gain for a travelling-wave tube is larger the lower the phase velocity of the wave guided by its effective impedance sheet. In the slipping-stream tube this wave has phase velocity about equal to the velocity of the electrons at the resonance layer and for $\Omega \gg \Omega_2$ the resonance layer is near the low-velocity edge with wave (3) and near the high-velocity edge with wave (2). Therefore the gain of wave (3) should be higher than the gain of wave (2).

The gains of waves (2) and (3) do not increase with frequency according to the law $\text{Im}(hV_0/\omega_0) \propto \omega^{1/3}$ of the travelling-wave tube. That they should not do so is apparent when it is remembered that the slow wave guided by the resonance layer will gather energy only from electrons which have velocities slightly greater than the phase velocity of the wave. Slower electrons will absorb energy from the wave. In a travelling-wave tube all the electrons travel slightly faster than the wave.

Having discussed the characteristics of the waves for $\alpha \approx 0.25$ in some detail let us now discuss the effect of changing α . Referring to Figure 2 we see that as α is increased from zero to 0.42 the branch point B of Figure 3 moves from $x = -0.715$ to $x = -1$. When $\alpha > 0.42$ the locus of x_a splits into two quite separate parts as shown in Figure 4. Waves (1) and (3) merge into one without any overlap in the gain-frequency characteristic of Figure 1. The examples $\alpha = 0.5$

and $\alpha = 0.8$ of Figure 1 illustrate this effect. They also show the steep rise in gain at high frequencies as $\alpha \rightarrow 1$, according to the law $\text{Im}(hV_0/\omega_0) \sim 0.061/(1-\alpha)$. Wave (2) however, starts at $\Omega_3 = (0.754/\alpha) - 0.247$ when x_a is at E. Between Ω_2 and $\Omega_3 = (0.754/\alpha) + 0.247$ it has a real propagation constant and when $\Omega > \Omega_2$ it has a gain which tends to the limiting value $\text{Im}(hV_0/\omega_0) \sim 0.061/(1+\alpha)$ as $\Omega \rightarrow \infty$.

That the behaviour of x_a and x_{-a} indicated in Figure 4 is not the whole story becomes evident when we consider what happens as α approaches the value unity. It is found that the loci of x_a and x_{-a} change in the manner shown in Figure 5. It now appears that the locus of x_a has a third branch GHKPQ (for clarity we have omitted the branch ECD from Figure 5). Moreover there is a critical value of α for which branches OAF and GHKPQ meet. This new branch passes through the singularity at $x = +1$ and runs along the real axis from $x = 1$ to $x = \infty$. When $\alpha > 0.946$, $\text{Re } x_a > \text{Re } x_{-a}$ along most of the curve GHKP; then it gives reverse waves, of which only the attenuating ones given by the negative sign of $\text{Im } h$ are of practical importance. Amplifying reverse waves are also given by this theory but they must always be accompanied by attenuating waves of larger amplitude.* When $\alpha = 0.946$ branches OAF and GHK meet at the point S, where $x_a = -0.30 + 1.98i$, and when $\alpha > 0.946$ they split into an upper branch FLG and a lower branch OMP. The corresponding loci of x_{-a} are indicated by dashed letters in Figure 5.

The most interesting feature of Figure 5 is the continued rise of the curve LF as $\alpha \rightarrow 1$. It can be shown that the curve LF tends to the asymptote

$$\text{Im } x_a = 0.061(1+\alpha)/(1-\alpha) \quad \text{as } x_{-a} \rightarrow 0.542 + 0.061i$$

and this gives the high gain, $\text{Im}(hV_0/\omega_0) = 0.061/(1-\alpha)$, as $\alpha \rightarrow 1$ at frequencies above ω_0 . It should not surprise one that the gain tends to infinity on this theory as $\alpha \rightarrow 1$. As the limit is approached both the velocity of the electrons at one edge of the beam and the velocity of the slow wave guided by the resonance layer tend to zero. The system should therefore behave as a travelling-wave tube with vanishing electron velocity which, in theory, would also have infinite gain.

It is worth while at this stage to compare the theoretical expressions for the maximum gain of the travelling-wave tube (TWT), two-beam tube (TBT) and slipping-stream tube (SST). We obtain the following values, expressed in decibels per unit length:

$$G_{\text{TWT}} \simeq 6 \left(\frac{\omega}{\omega_0} \right)^{1/3} \frac{\omega_0}{\bar{V}},$$

$$G_{\text{TBT}} \simeq 4.3 \frac{\omega_0}{\bar{V}}, \quad \text{where } \bar{V} \text{ is the mean velocity of the two electron streams,}$$

$$G_{\text{SST}} \simeq 2.1 \frac{\omega_0}{\bar{V}} \quad \text{for the TBT-mode}$$

and

$$G_{\text{SST}} \simeq 0.53 \frac{\omega_0}{V_{-a}} \quad \text{for the TWT-mode.}$$

* A discussion of reverse waves will be found in a forthcoming paper by G. G. Macfarlane and Mrs. A. M. Woodward.

These expressions apply to ideal systems in which no metal and other losses occur and in which the electrons move in regular streams without any random thermal motions.

Maximum amplification is obtained in all cases when the electron velocity is a minimum. It is therefore reasonable to compare the systems for the same minimum velocity. Moreover, in order to obtain gain at frequencies above ω_0 with the two-beam tube it is necessary for V_{-a} to be nearly equal to \bar{V} . Therefore for comparison of the performance of these ideal tubes as amplifiers for frequencies much higher than ω_0 we can put $V = \bar{V} = V_{-a}$. Then we get

$$G_{\text{TWT}} : G_{\text{TBT}} : G_{\text{SST}}(\text{TBT-mode}) : G_{\text{SST}}(\text{TWT-mode}) \\ = 6(\omega/\omega_0)^{1/3} : 4.3 : 2.1 : 0.53.$$

A more detailed discussion of the effects of losses, thermal motion of the electrons, and possible feeding systems is beyond the scope of this paper.

It still remains to mention another result which may be important in consideration of the use to which a slipping-stream tube might be put. In Figure 5 it is shown that the portion of the real axis from $x=1$ to $x=\infty$ is a common locus for x_a and x_{-a} for all values of α . There is no gain or attenuation for these waves and they are reverse uniform waves. When α is equal to unity the phase velocity depends on frequency for $\Omega > 1$ as shown in Figure 7. These reverse waves may

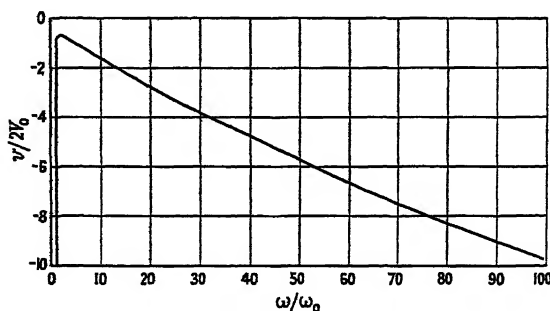


Figure 7. The phase velocity of the reverse uniform waves as a function of frequency, $\alpha=1$.

constitute a serious drawback to the use of a slipping-stream tube as an amplifier unless it can be found how to avoid exciting them. The fact, however, that Haeff has obtained a stable amplifier using a slipping stream is encouraging.

Summing up the discussion we can say that a slipping-stream tube will behave differently according to the fractional velocity range of the electrons: if

$$(V_a - V_{-a})/(V_a + V_{-a}) = \alpha < 0.42$$

it behaves primarily like a two-beam tube with cut-off frequency $\Omega_1 = 1/\alpha\sqrt{2}$ but it has only about one half its maximum gain, viz. $2.1\omega_0/V_0$ decibels per unit length. In addition it has a gain of $0.53\omega_0/V_{-a}$ for $\Omega \gg \Omega_1$. However, if $\alpha > 0.42$ it behaves primarily like a two-wave tube with a gain of $0.53\omega_0/V_{-a}$ decibels per unit length. Comparison with the travelling-wave tube and two-beam tube on the basis of the same minimum electron velocity shows that the slipping-stream tube has a lower maximum rate of gain than either of the other two.

(iii) *Properties of the Amplifying Wave in a Slipping-stream Travelling-wave Tube*

In this section we give results for a slipping stream inside a waveguide. The waveguide is such that a slow TM-wave of phase velocity v_0 can travel along it when there is no electron beam inside it. The velocity is taken to vary from 0 to $2V_0$ ($\alpha=1$). The system is a form of travelling-wave tube using a slipping stream of electrons instead of a uniform stream.

The theory, by means of which the following results were obtained, is set out in §2(iv). It was shown in that section that the waveguide could be defined in terms of an impedance sheet. This concept obviates the necessity of specifying the physical form of loaded guide required to support a slow wave of TM-type. It effects at once a unification and simplification because a TM-wave of phase velocity v_0 with the field distribution given in §2(iv) can be guided by different physical structures. The object of our analysis is to discover how the propagation constant h varies with frequency for a range of values of v_0 in the neighbourhood of $2V_0$. In most physical structures used to obtain such slow waves the phase velocity v_0 will vary with frequency. By giving a map of h marked in contours of constant $2V_0/v_0$ and ω/ω_0 the dependence of the propagation constant on frequency can be quickly determined when the variation of v_0 with frequency is known. Such a contour map is shown in Figure 8, in which, instead of h , the quantity

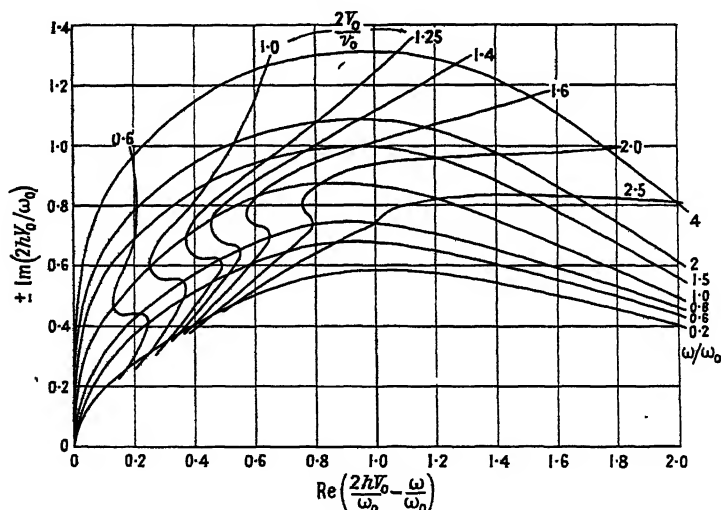


Figure 8. Contour map of the propagation constant for the slipping stream travelling-wave tube v_0 is the phase velocity of the TM-wave in the empty guide.

given is $x_a = (\omega/\omega_0) - (2hV_0/\omega_0)$. Since $\alpha=1$, $x_a = \omega/\omega_0$, so that the real component of h can be calculated immediately from the curves. The ordinate is simply $\text{Im}(2hV_0/\omega_0)$. From the figure it is therefore clear that the maximum rate of gain is obtained when $v_0 \simeq 2V_0$, that is when the phase velocity in the empty guide is about equal to the velocity of the fastest electrons. It is interesting to compare these results with the corresponding results for a travelling-wave tube with uniform beam, which are shown in Figure 9 (Woodward, unpublished work).

In this figure the velocity of the electrons is V and the phase velocity in the empty guide is v . The similarity of the characteristics at any rate for $\omega > \omega_0$ is immediately evident. The maximum gain is slightly greater for the slipping-stream tube. There is, however, one noticeable difference. When $v_0 < V$ there is an

upper limit to the gain with the travelling-wave tube, but when $v_0 < 2V_0$ there is no upper limit with the loaded slipping-stream tube. The reason for this is that the electron velocity ranges in value from 0 to $2V_0$ in the slipping-stream tube so that there are always some electrons for which the condition, $v_0 \geq V$, for gain at the highest frequencies is satisfied. However, for $v_0 > 2V_0$ the gain falls to low

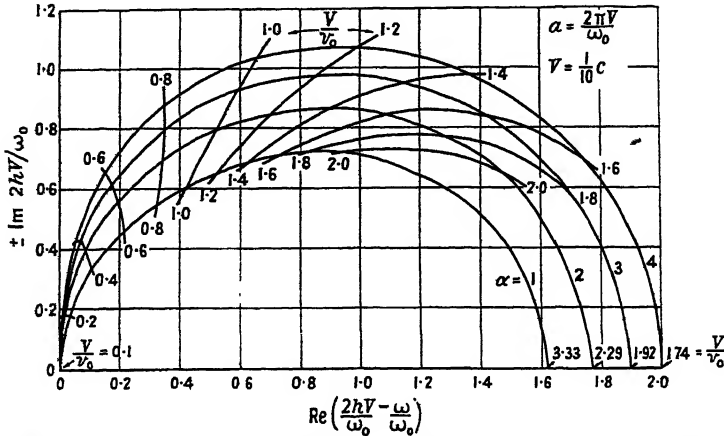


Figure 9. Contour map of the propagation constant for a uniform stream travelling-wave tube.

values at high frequencies because the electrons which can give up energy to the waves are further from the impedance sheet. Moreover the field falls off rapidly from the impedance sheet.

In Figure 10 we have plotted the phase velocity of the wave as a function of frequency. It is always less than the velocity of the fastest electrons.

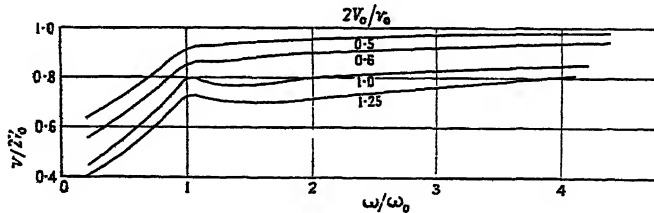


Figure 10. The phase velocity v as a function of frequency for the slipping stream travelling-wave tube. The parameter is $2V_0/v_0$, where v_0 is the phase velocity in the empty guide.

§ 4. ADDITIONAL NOTES

(a) The non-oscillatory slipping stream requires applied electric and magnetic fields to ensure parallel motion of the electrons. From (3) the magnetic flux density is

$$B_x = \frac{m}{e} \omega_0 = 5.68 \times 10^{-8} \omega_0 \text{ gauss.}$$

From (1) the potential in the beam is $\phi = mV^2/2e$. If the electric field is maintained by a potential difference between two parallel plates at $y = \pm b$ the potential on the plates must be

$$\phi(b) = \frac{m}{2e} V_2^2 + (b-a) \frac{m\omega_0}{e} V_2; \quad \phi(-b) = \frac{m}{2e} V_1^2 - (b-a) \frac{m\omega_0}{e} V_1,$$

where $\phi = 0$ is the potential of the cathode from which the electrons are emitted

(b) In order to study the basic phenomena we have described a planar form of slipping-stream tube. Other forms can be thought of which may be of greater practicability. For example an annular beam might be used between the inner and outer conductors of a coaxial line; the magnetic field, which would be azimuthal, could be produced by axial currents in the inner and outer conductors.

(c) From the small-signal theory it is not possible to deduce the behaviour of the system at large signal amplitudes. Nevertheless it is interesting to try to discuss a few general points connected with the transformation of energy in the system.

It can be easily shown that the mean flow of energy across the section of the tube is

$$W = - \int_{\Gamma} (\mathbf{E} \times \mathbf{H}) \cdot \mathbf{n} da - \frac{m}{2e} \int (\rho V^2) \mathbf{V} \cdot \mathbf{n} da. \quad \dots\dots(43)$$

This latter integral is the mean rate at which kinetic energy of the electrons is passing the section Γ .

At the input section the R.F. energy can be neglected, $V = V_0 + \omega_0 y$, and Γ extends from $y = -a$ to $y = +a$. Evaluating the integral we get

$$W = \frac{m}{e} \rho a (V_0^2 + a^2 \omega_0^2) V_0. \quad \dots\dots(44)$$

At a section farther along the tube this power is partly converted into oscillatory power. If we still think of the beam in two parts, one D.C. and the other A.C., we see that the A.C. power can increase at the expense of the D.C. power either by a decrease in mean velocity or a decrease of density of the electrons or by both. Now in a travelling-wave tube with strong axial magnetic field the D.C. power comes primarily from a decrease in mean velocity because lateral spreading is inhibited, but in the slipping-stream tube the beam is at liberty to spread and lose average density as well as mean velocity. On this account it seems likely that there will be a more efficient transfer of power from D.C. to A.C. at large amplitude with the slipping-stream tube.

ACKNOWLEDGMENT

The work described above has been carried out for the Department of Scientific and Industrial Research and is published by permission of the Director of Radio Research, D.S.I.R., the Chief Scientist, Ministry of Supply, and the Controller of H.M. Stationery Office.

REFERENCES

- BUNEMANN, O., 1944, *C. V. D. Report Mag.*, 37. A useful account of this work is given by L. R. Walker in *Microwave Magnetrons*, No. 6 of *Radiation Laboratory Series* (McGraw-Hill Book Co.), 1948, pp. 253-264.
- GABOR, D., 1945, *Proc. Inst. Radio Engrs.*, N. Y., 33, 804.
- HAEFF, A. V., 1949, *Proc. Inst. Radio Engrs.*, N. Y., 37, 4.
- HAHN, W. C., 1939, *Gen. Elect. Rev.*, 42, 258.
- RAMO, S., 1939, *Phys. Rev.*, 56, 276.
- RAMSEY, A. S., 1935, *A Treatise on Hydrodynamics*, Pt. II, 4th edn. (London: G. Bell and Son), p. 13.
- SMITH, L. P., and HARTMAN, P. L., 1940, *J. Appl. Phys.*, 11, 220.

APPENDIX I

DERIVATION OF THE HAMILTON-JACOBI EQUATION

The following discussion is restricted to non-relativistic motion.

In terms of the scalar potential ϕ and the vector potential \mathbf{A} the equation of motion of an electron in Newtonian form is

$$m \frac{d\mathbf{V}}{dt} = -e\mathbf{E} - e\mathbf{V} \wedge \mathbf{B} = e \left(\frac{\partial \mathbf{A}}{\partial t} + \text{grad } \phi - \mathbf{V} \wedge \text{curl } \mathbf{A} \right). \quad \dots (45)$$

For electrons moving in a stream the left hand side of (45) can be transformed, as in classical hydrodynamics, to give

$$\frac{d\mathbf{V}}{dt} = \frac{\partial \mathbf{V}}{\partial t} + (\mathbf{V} \cdot \text{grad})\mathbf{V} = \frac{\partial \mathbf{V}}{\partial t} + \frac{1}{2} \text{grad } V^2 - \mathbf{V} \wedge \text{curl } \mathbf{V}. \quad \dots (46)$$

Then (45) becomes

$$\frac{\partial}{\partial t} (m\mathbf{V} - e\mathbf{A}) = \mathbf{V} \wedge \text{curl } (m\mathbf{V} - e\mathbf{A}) + \text{grad } (e\phi - \frac{1}{2}mV^2). \quad \dots (47)$$

If the motion is such that

$$\mathbf{V} \wedge \text{curl } (m\mathbf{V} - e\mathbf{A}) = 0, \quad \dots (48)$$

$m\mathbf{V} - e\mathbf{A}$ must be the gradient of some scalar, ψ say. Then

$$m\mathbf{V} - e\mathbf{A} = \text{grad } \psi \quad \dots (49)$$

and, on integrating (47),

$$\frac{\partial \psi}{\partial t} + \frac{1}{2}mV^2 - e\phi = 0, \quad \dots (50)$$

since the constant of integration, which may be a function of time, can always be reduced to zero by suitable choice of ψ .

Eliminating \mathbf{V} from (49) and (50) gives

$$\frac{\partial \psi}{\partial t} + \frac{1}{2}m(e\mathbf{A} + \text{grad } \psi)^2 = e\phi,$$

which is the Hamilton-Jacobi equation in §2(i).

It can be shown (Gabor 1945) that the condition $\text{curl } m\mathbf{V} = e\mathbf{B}$ is satisfied at all points of the motion if the electrons leave the cathode with zero velocity and there is no component of magnetic field normal to the cathode surface.

Given that $\text{curl } m\mathbf{V} = e\mathbf{B}$, the two-dimensional problem treated in this paper demands that the magnetic field maintaining the velocity distributions be at right angles to the plane of the motion, and, in addition, be of a specified magnitude.

For if $\mathbf{V} = [0, V_y, V_z]$ and \mathbf{V} is independent of x , then

$$\text{curl } \mathbf{V} = \left[\frac{\partial V_z}{\partial y} - \frac{\partial V_y}{\partial z}, 0, 0 \right],$$

so that $\mathbf{B} = [B_x, 0, 0]$, where

$$B_x = \frac{m}{e} \left(\frac{\partial V_z}{\partial y} - \frac{\partial V_y}{\partial z} \right).$$

APPENDIX II

If the equation of the boundary surface is

$$F(y, z, t) = y - a - \Delta e^{i\omega t - ihz} = 0,$$

the function F must satisfy the equation (Ramsey.1935)

$$\frac{\partial E}{\partial t} + V_y \frac{\partial F}{\partial y} + V_z \frac{\partial F}{\partial z} = 0.$$

Since $V_z = V$ and $V_y = \frac{1}{m} \frac{\partial \psi_1}{\partial y} e^{i\omega t - ihz}$, this gives

$$\Delta = \frac{1}{m} \frac{\partial \psi_1}{\partial y} / i(\omega - Vh).$$

Theory of the Production of an Ionized Layer in a Non-Isothermal Atmosphere Neglecting the Earth's Curvature, and its Application to Experimental Results

By J. A. GLEDHILL AND M. E. SZENDREI

Rhodes University College, Grahamstown, South Africa

Communicated by M. Blackman; MS. received 31st March 1949, and in amended form 26th January 1950

ABSTRACT. A new theory of ionospheric layer formation is developed, in which the temperature is assumed to vary linearly with height. The equations are compared at each step with those obtained by Chapman in his theory of layer formation in an isothermal atmosphere. The equations for the maximum of electron density and its height are also given. The effect of the parameters on the shape of the layer is shown in graphical form.

The equations are somewhat complex in form, but an ingenious graphical method has been devised suitable for the application of the theory to results given in the form of a Booker and Seaton parabolic distribution of electronic density with height. By application of the theory to mean hourly values for four summer months in South Africa, values are obtained for the temperature gradient, the temperature at 200 km., and its variation over the middle part of the day. The results obtained are in accordance with previous estimates and offer numerical confirmation of the theory that the atmosphere expands bodily upwards during the middle part of a summer day.

§ 1. THE THEORY

1 (i). Introduction

THE most prominent among the theories put forward to account for the formation of ionospheric layers is that of Chapman (1931, 1939), which attributes their production to the ionizing action of solar ultra-violet radiation on the earth's atmosphere. Other possible causes, such as corpuscular radiation, cosmic rays, alpha and beta rays from various sources, meteors, etc., are entirely neglected; owing to the irregular nature of most of them, they are

intractable to theoretical treatment at present. In Chapman's theory a large number of simplifying assumptions is necessarily made:

- (i) the radiation is monochromatic;
- (ii) the density of the atmosphere, on which the radiation is incident, varies exponentially with height;
- (iii) the atmosphere is at rest and uniform in composition;
- (iv) the temperature is the same at all points in the atmosphere at all times;
- (v) the force of gravitation is constant with height;
- (vi) all the absorbed energy causes ionization.

Charts showing the ionization of the upper atmosphere have been constructed by Millington (1932, 1935), using Chapman's theory, but do not seem to have had much application.

Various approximate theories have been put forward by Hulburt (1928, 1938); finally an attempt was made (Hulburt 1939a) to fit a theoretical curve to the observed mean values of maximum electron density for the E layer.

Comparison shows that Chapman's and Hulburt's theories are the same in fundamental content, being based on the same initial assumptions. Their application to the E region has met with considerable success, but few attempts have been made to apply them to the higher more irregular regions.

To develop a theory which will represent more nearly existing conditions in the F regions, it seems advisable to reject Chapman's assumption of an isothermal atmosphere and, instead, to allow the temperature to vary with height in a known way; as the simplest variation possible, it has been assumed in this paper that the temperature varies linearly with height from a given level upwards. The distribution of molecules with height in such an atmosphere will no longer be exponential and will first be derived.

1 (ii). *Distribution of Molecules with Height*

Assuming no vertical movement of the atmosphere, the increase in pressure dp corresponding to an increase in height dh is given by

$$-dp = g\rho dh, \quad \dots\dots(1)$$

where g is the acceleration due to gravity and ρ the density at height h . We use the simple gas laws, and put n' for the number of molecules per cm^3 at height h and n'_0 for the corresponding value at a reference level h_0 , the temperature at this height being T_0 . On substitution and integration between the limits n'_0 and n' , corresponding to heights h_0 and h , we have

$$n' = n'_0 \frac{T_0}{T} \exp\left(-\frac{mg}{k} \int_{h_0}^h \frac{dh}{T}\right), \quad \dots\dots(2)$$

where k = Boltzmann's constant, and m = mean molecular mass.

In this paper we limit the discussion to the case where T is a linear function of height and is given by

$$T = T_0 + \gamma(h - h_0). \quad \dots\dots(3)$$

Substitution in equation (2) and integration then yields

$$n' = n'_0 \left\{ 1 + \gamma \frac{h - h_0}{T_0} \right\}^{-(1 + G/\gamma)}, \quad \dots\dots(4)$$

where we have written

$$C = \frac{mg}{k}. \quad \dots\dots(5)$$

On expressing this as a binomial expansion and inserting the limit $\gamma=0$, equation (4) reduces to

$$n' = n'_0 \exp \left\{ -\frac{C}{T_0} (h - h_0) \right\}, \quad \dots\dots(6)$$

which is the exponential distribution assumed by Chapman and Hulburt.

Table 1. Values of n'/n'_0 for various γ 's

h (km.)	$\gamma=0.00$ (deg/km.)	$\gamma=2.02$ (deg/km.)	$\gamma=4.05$ (deg/km.)	$\gamma=8.10$ (deg/km.)
200	1.00	1.00	1.00	1.00
225	3.64×10^{-1}	3.44×10^{-1}	3.25×10^{-1}	2.92×10^{-1}
250	1.33×10^{-1}	1.32×10^{-1}	1.30×10^{-1}	1.22×10^{-1}
275	4.78×10^{-2}	5.54×10^{-2}	5.98×10^{-2}	6.25×10^{-2}
300	1.74×10^{-2}	2.53×10^{-2}	3.05×10^{-2}	3.64×10^{-2}
325	6.28×10^{-3}	1.22×10^{-2}	1.69×10^{-2}	2.27×10^{-2}
350	2.29×10^{-3}	6.24×10^{-3}	9.93×10^{-3}	1.52×10^{-2}

The effect of a temperature gradient γ on the molecular density may be seen from Table 1, which has been worked out assuming the values $h_0=200$ km., $T_0=400^\circ\text{K.}$, $g=900$ cm.sec⁻², $m=15 \times 1.66 \times 10^{-24}$ gm. and $k=1.38 \times 10^{-16}$ erg. deg⁻¹.

1 (iii). Intensity of Incident Radiation

The intensity I of the incident radiation (assumed monochromatic) will decrease as it passes through the atmosphere. Suppose that the intensity is increased by an amount dI in increasing the height from h to $h+dh$. If χ , the solar zenith distance, is small, so that we may consider the earth's surface to be flat,

$$dI = \beta n' I \sec \chi \cdot dh, \quad \dots\dots(7)$$

where β is the atomic absorption coefficient. Substitution from equation (2), and subsequent integration yields

$$I = I_\infty \exp \left\{ n'_0 T_0 \beta \sec \chi \int_\infty^h \frac{1}{T} \exp \left(-\frac{mg}{k} \int_{h_0}^h \frac{dh}{T} \right) dh \right\}, \quad \dots\dots(8)$$

in which I_∞ is the intensity of the beam before it has reached the earth's atmosphere and I the intensity at height h .

1 (iv). Rate of Ion Production

Suppose q is the number of ion pairs produced per cm³ per second at height h . Then if w is the energy absorbed in ionizing one molecule, and assuming that all the energy absorbed produces ionization,

$$q = \frac{\beta n' I}{w}. \quad \dots\dots(9)$$

Hence, on substituting from (8),

$$q = \frac{\beta n'_0 I_\infty}{w} \cdot \frac{T_0}{T} \exp \left\{ -C \int_{h_0}^h \frac{dh}{T} + n'_0 T_0 \beta \sec \chi \int_{\infty}^h \frac{1}{T} \exp \left(-C \int_{h_0}^h \frac{dh}{T} \right) dh \right\}. \quad \dots\dots(10)$$

With a linear temperature gradient, according to equation (3), this becomes

$$\begin{aligned} q &= q_b \\ &= \frac{\beta n'_0 I_\infty}{w} \left(1 + \gamma \frac{h - h_0}{T_0} \right)^{-(1+C/\gamma)} \exp \left\{ -\frac{n'_0 \beta T_0}{C} \sec \chi \left(1 + \gamma \frac{h - h_0}{T_0} \right)^{-C/\gamma} \right\}. \end{aligned} \quad \dots\dots(11)$$

On proceeding to the limit $\gamma = 0$, as before, this reduces to

$$\begin{aligned} q &= q_a \\ &= \frac{\beta n'_0 I_\infty}{w} \exp \left\{ -\frac{C}{T_0} (h - h_0) - \frac{n'_0 \beta T_0}{C} \sec \chi \exp \left[-\frac{C}{T_0} (h - h_0) \right] \right\} \end{aligned} \quad \dots\dots(12)$$

which, in the present notation, agrees with Hulburt and with Chapman.

1(v). *Layer Formation*

If ions are produced according to equation (11) or (12), there will be an equal number n of positive ions and electrons per cm^3 . The rate of increase of ion density is then given by the well-known equation

$$\frac{dn}{dt} = q - \alpha n^2, \quad \dots\dots(13)$$

in which α represents the effective recombination coefficient.

Case (a).—It is instructive to consider first the solution of equation (13) for the constant temperature case. Substituting from equation (12),

$$\frac{dn}{dt} + \alpha n^2 = \frac{\beta n'_0 I_\infty}{w} \exp \left\{ -\frac{C}{T_0} (h - h_0) - \frac{n'_0 \beta T_0}{C} \sec \chi \exp \left[-\frac{C}{T_0} (h - h_0) \right] \right\}. \quad \dots\dots(14)$$

According to Chapman, this equation cannot be solved explicitly. Hulburt, however, has pointed out that during the middle part of the day dn/dt is small, and may be assumed to be zero to obtain an approximate equation. This assumption is confirmed by the experimental results to be described in § 2 of this paper. Equation (14) then reduces to

$$n_a^2 = B \exp \left\{ -\frac{C}{T_0} (h - h_0) - \frac{1}{F} \exp \left[-\frac{C}{T_0} (h - h_0) \right] \right\}, \quad \dots\dots(15)$$

where we have written

$$B = \beta n'_0 I_\infty / \alpha w \quad \dots\dots(16)$$

and

$$F = C \cos \chi / n'_0 \beta T_0. \quad \dots\dots(17)$$

Case (b).—A linear temperature gradient, given by equation (3), will now be considered. As the form of the equation is more complex than in case (a), we proceed at once to make the approximation $dn/dt=0$, obtaining

$$n_b^2 = B \left(1 + \gamma \frac{h-h_0}{T_0} \right)^{-(1+C/\gamma)} \exp \left\{ -\frac{1}{F} \left(1 + \gamma \frac{h-h_0}{T_0} \right)^{-C/\gamma} \right\}. \quad \dots\dots (18)$$

Equations (15) and (18) give the value of the electron density at any point in the layer.

1 (vi). *Maximum of Electron Density*

The maximum-with-height value N of the electron density, n , may be found by equating dn/dh to zero. When this is done with equations (15) and (18) in turn, we find that the required conditions are respectively

$$\frac{1}{F} \exp \left\{ -\frac{C}{T_0} (h-h_0) \right\} = 1 \quad \dots\dots (19)$$

and

$$\frac{1}{F} \left(1 + \gamma \frac{h-h_0}{T_0} \right)^{-C/\gamma} = 1 + \frac{\gamma}{C}. \quad \dots\dots (20)$$

When substituted back in the original equations, we have

$$N_a^2 = BF \exp(-1) \quad \dots\dots (21)$$

and

$$N_b^2 = B \left\{ F \left(1 + \frac{\gamma}{C} \right) \right\}^{(1+\gamma/C)} \exp \left\{ - \left(1 + \frac{\gamma}{C} \right) \right\}, \quad \dots\dots (22)$$

giving the required maxima for the two cases.

The height of maximum electron density, h_M , is given by equations (19) and (20), which, for future reference, are now written as

$$(h_M)_a = h_0 - \frac{T_0}{C} \ln F, \quad \dots\dots (23)$$

$$(h_M)_b = h_0 + \frac{T_0}{\gamma} \left[\left\{ F \left(1 + \frac{\gamma}{C} \right) \right\}^{-\gamma/C} - 1 \right]. \quad \dots\dots (24)$$

1 (vii). *Comparison of the Two Cases*

The changes in behaviour brought about by the assumption of a linear temperature gradient with height may be illustrated by means of the ratios

$$\frac{N_b}{N_a} \quad \text{and} \quad \frac{(h_M)_b - h_0}{(h_M)_a - h_0}.$$

By this method the use of a number of questionable constants may be eliminated. The theory assumes that the composition of the atmosphere at the reference height h_0 should be the same as that throughout the layer. There is reason to believe that the atmospheric gases exist mainly in atomic form from 200 km. upwards (Hulburt 1939 b); h_0 has therefore been chosen as 200 km., because the value of T_0 becomes less certain as the height increases. Further, putting $m=15$ atomic weight units, $g=900$ cm.sec⁻² and $k=1.38 \times 10^{-16}$ erg.deg⁻¹, the value of C becomes 16.2 deg.km⁻¹. Finally, for purposes of this comparison only, assuming $T_0=360^\circ$ K., $n_0=10^{11}$ cm⁻³ and $\beta=10^{-17}$ cm², on very debatable grounds, and putting $\chi=0$ for simplicity, we find $F=4.4 \times 10^{-2}$.

Dividing equation (22) by (21), we obtain

$$\left(\frac{N_b}{N_a}\right)^2 = F^{\gamma/C} \cdot \left(1 + \frac{\gamma}{C}\right)^{(1+\gamma/C)} \cdot \exp(-\gamma/C), \quad \dots\dots(25)$$

and from equations (23) and (24)

$$\frac{(h_M)_b - h_0}{(h_M)_a - h_0} = -\frac{C}{\gamma} \frac{F \left(1 + \frac{\gamma}{C}\right)^{-\gamma/C} - 1}{\ln F}. \quad \dots\dots(26)$$

To obtain some idea of the absolute difference in height of the maxima, we use

$$(h_M)_b - (h_M)_a = T_0 \left[\frac{1}{\gamma} \left\{ \left[F \left(1 + \frac{\gamma}{C}\right) \right]^{-\gamma/C} - 1 \right\} + \frac{1}{C} \ln F \right], \quad \dots\dots(27)$$

obtained by subtracting equation (23) from equation (24).

The values obtained are shown graphically in Figures 1 and 2. These graphs

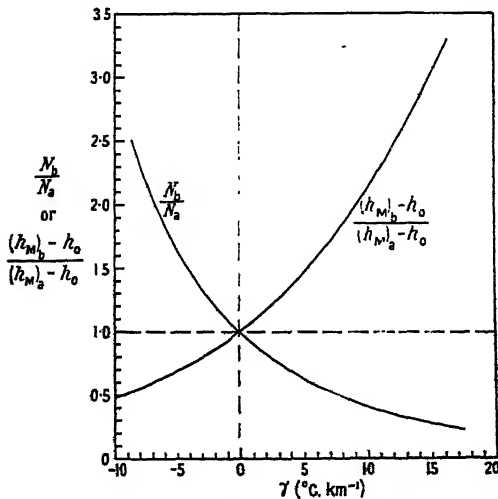


Figure 1.

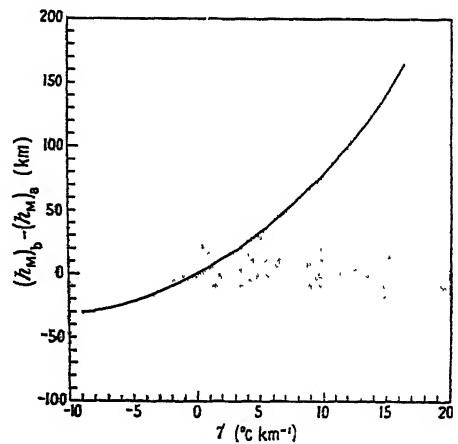


Figure 2.

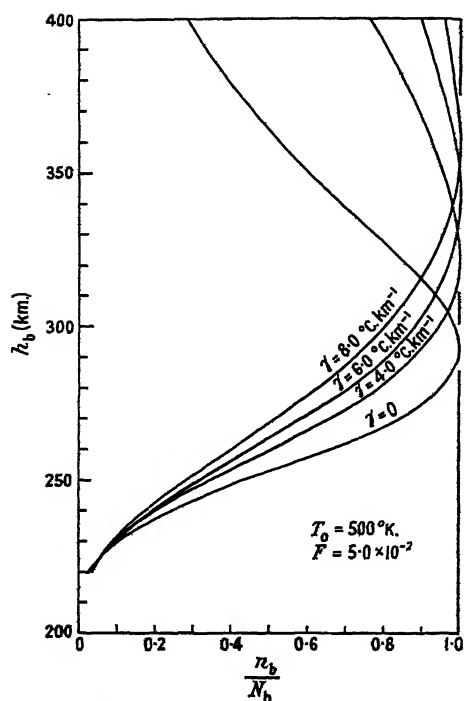
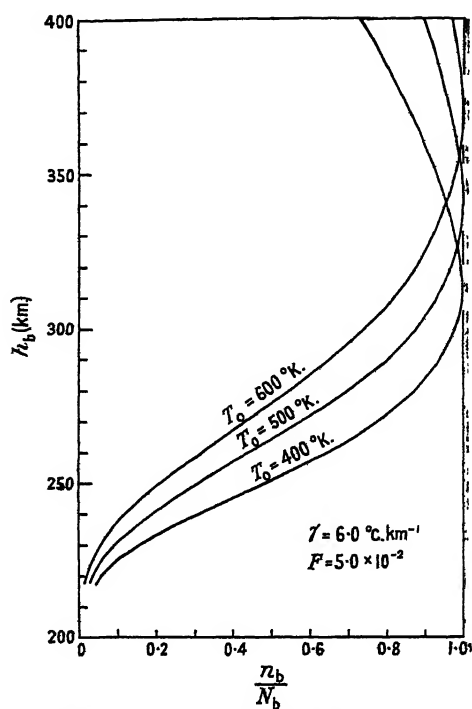
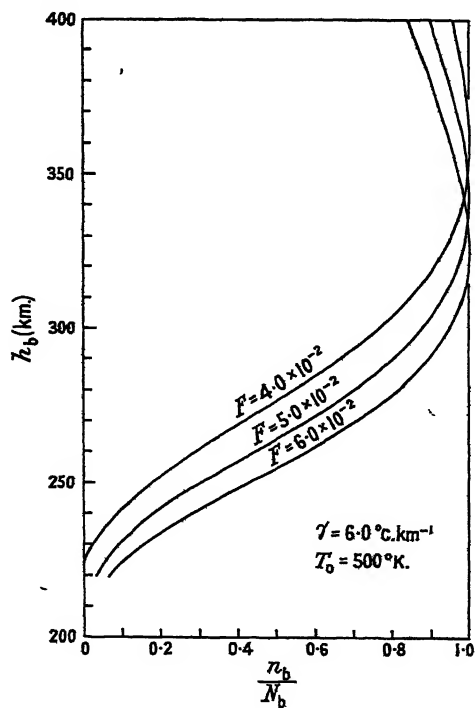
show that a positive value of γ produces a lowering of the maximum electron density in the layer and raises its height.

1 (viii). Effect of Variables on the Shape of the Layer

As the general theory, summarized in equations (18), (22) and (24), does not lend itself to easy physical interpretation, it has been thought advisable to represent graphically the variations in shape of the layers brought about by changes in the variables. In order to eliminate as far as possible numerical errors arising from the use of uncertain values of the constants, equations (18) and (22) may be divided, giving

$$\begin{aligned} \left(\frac{n_b}{N_b}\right)^2 &= \left(1 + \gamma \frac{h_b - h_0}{T_0}\right)^{-(1+C/\gamma)} \left\{ F \left(1 + \frac{\gamma}{C}\right) \right\}^{-(1+\gamma/C)} \\ &\times \exp \left\{ \left(1 + \frac{\gamma}{C}\right) - \frac{1}{F} \left(1 + \gamma \frac{h_b - h_0}{T_0}\right)^{-C/\gamma} \right\} \quad \dots\dots(28) \end{aligned}$$

in which B does not appear. On the right-hand side of equation (28) we are thus left with four parameters, the height h_b , the temperature T_0 at 200 km., the temperature gradient γ , and F . In Figures 3, 4 and 5, n_b/N_b is plotted against h_b for various values of γ , T_0 and F , using equation (28).


 Figure 3. Effect of γ on shape of layer.

 Figure 4. Effect of T_0 on shape of layer.

 Figure 5. Effect of F on shape of layer.

In Figure 3, the curve for $\gamma=0$ is drawn according to Chapman's theory, and represents the fractional distribution of electrons at constant temperature. We note that a positive γ raises the height h_M . In addition, it is seen to increase the thickness of the layer, and the electron density above the maximum decreases more slowly than in an isothermal atmosphere. The importance of a knowledge of true layer thickness in any attempt to estimate γ is thus clearly brought out.

From Figure 4, it will be seen that an increase in T_0 produces an increase in the height of maximum electron density, but the thickness of the layer is not altered to any great extent. In Figure 5, the thickness of the layer is again practically unchanged by a change in F , but the height h_M decreases with increase in F . Remembering that for a given T_0 (and therefore n_0) F is proportional to $\cos \chi$ from equation (17), these curves show a result analogous to the theory for an isothermal atmosphere, i.e. the height of maximum electron density decreases to a minimum at local noon, all the other parameters remaining constant. However, since the theory of expansion demands atmospheric heating during the middle part of the day, T_0 and γ cannot be taken as independent of time. If the increase in T_0 is greater than that of $\cos \chi$, it alone would account for a rise in h_M , even if γ remained unchanged.

§ 2. EXPERIMENTAL DATA

Before developing a method by which the theory just expounded may be applied to experimental results, it is convenient to consider how many and which quantities may be derived from ionospheric measurements. Records of ionospheric conditions are almost universally given in the form of graphs of virtual height plotted against frequency; these are obtained either by automatic equipment such as that developed by Berkner, Wells and Seaton (1936) at the Carnegie Institute, or by a more modest form of manually operated equipment, where financial resources are not readily available or the proposed experimental work is not of a very permanent nature. From such records may be obtained:

- (i) the maximum electron density N in the layer, which is directly proportional to the square of the critical frequency for the ordinary ray;
- (ii) the true height, h_M , of maximum density, and semi-thickness τ of the layer, which may be obtained without much difficulty either by the method developed by Booker and Seaton (1940) or by a more recent method developed by Pierce (1947).

At this stage it is interesting to remark how investigators have, up to the present, insisted on jumping to conclusions, basing their reasoning entirely on critical frequencies and virtual heights. In dealing with a problem as complex as the ionosphere, it is an obvious advantage to rely as far as possible on experimental results and thus to eliminate, in part at least, some of the little known parameters.

The photographic records used for purposes of testing this theory and gaining some information of conditions in the F_2 region were taken in this laboratory with manually operated equipment assembled by the writers during the war. A pulse transmitter of continuously variable frequency covered a range from 1.8 to 13.9 Mc/s.; its mean power output was about 200 watts during pulses. Records were made on 35 mm. film and scaled through a microfilm reader giving an approximate magnification of 10. By the help of frequency marks made on

the record at regular intervals of 0.2 or 0.5 Mc/s. by means of a height calibrating oscillator working at 1,500 c/s., frequencies could be read to 0.01 Mc/s. and heights could be reproduced to 2 km.

A series of photographic records was taken, one on each hour on alternate days between 1st November 1945 and 28th February 1946, the summer season

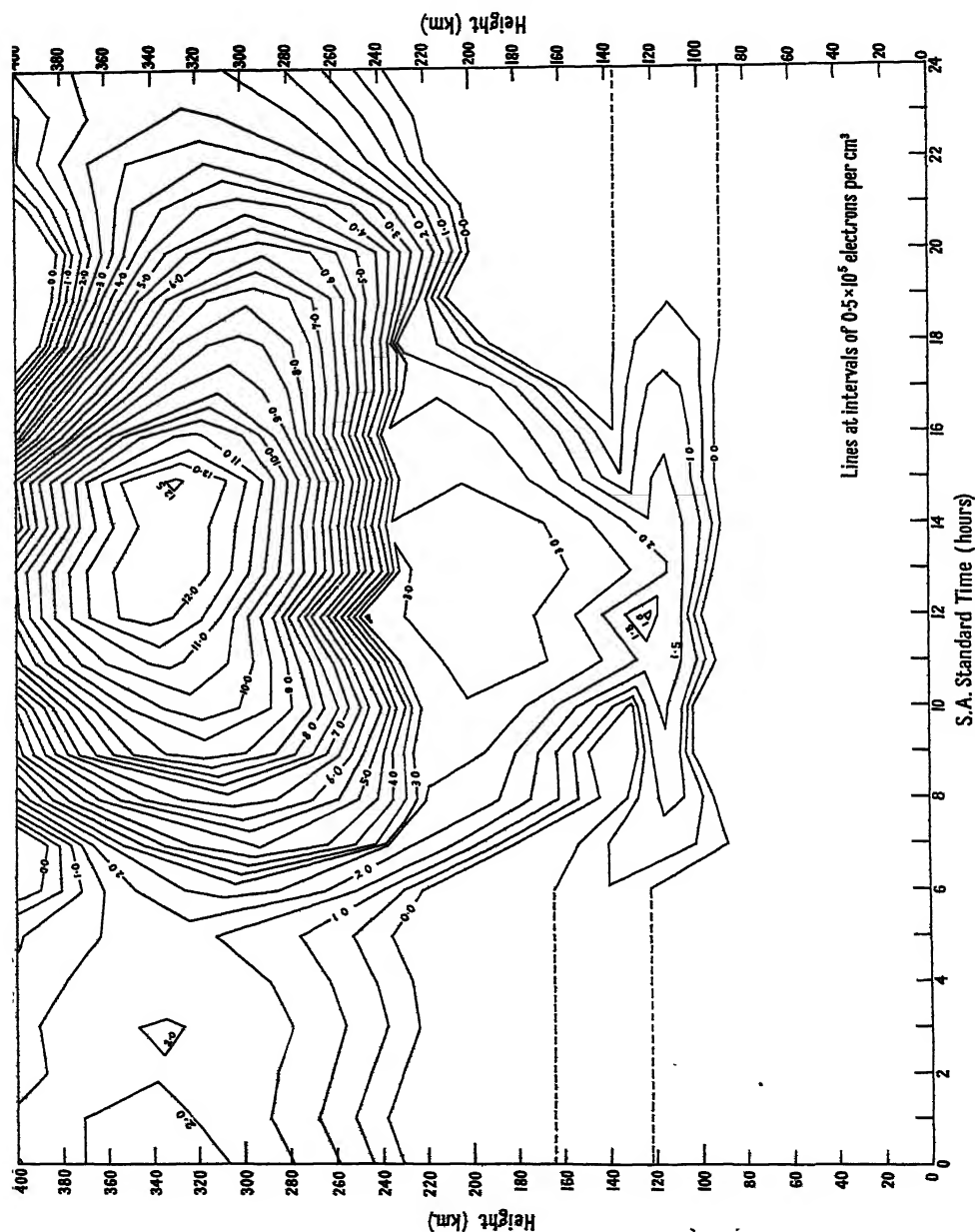


Figure 6. Electron density map. February 1946

in South Africa; this choice of intervals was dictated by the shortage of observing staff available. All the records were completely scaled according to the method of Booker and Seaton (1940); heights were read at the 8 frequencies suggested by Booker and Seaton, and coupled in four pairs, each of them giving the true

height of maximum density, h_M , and semi-thickness, τ , of the layer. The mean of the four values was calculated for each quantity. Where necessary, corrections due to retardation by lower layers were applied (Booker and Seaton 1940). Since a significant correlation was found between magnetically disturbed and ionospherically disturbed days, all days with magnetic character figure 2 were eliminated in taking the monthly mean values of N , h_M and τ . As some large deviations still remained, the more outstanding ones were removed, using the well established test that deviations larger than three times the probable deviation are unlikely to be due to chance alone. With the adjusted monthly mean values for N , h_M and τ thus obtained, it was possible to draw the Booker and Seaton parabolic distribution of (electron density, true height) for each hour of the average day for a given month.

It has been found convenient to represent all the information about the behaviour of the ionosphere on the average day for each month by a graph showing lines of constant n on a coordinate plane of time and height. This plot is called an 'electron density map': the map for the February data is shown in Figure 6 as an example. It serves to illustrate the important point that dn/dt is in fact nearly zero during the daytime in parts of the F_2 layer below the electron density maximum.

§ 3. APPLICATION OF THE THEORY TO EXPERIMENTAL RESULTS

3 (i). *The Method of applying the Theory*

The application to experimental results of a theory with equations as complex as those derived in § 1 of this paper is a problem which must present some difficulty. From the theory there emerge the three useful equations (18), (22) and (24). Considering C as known, we see that these equations contain no less than four unknowns: B , γ , T_0 and F . B may be eliminated by dividing (18) by (22), thus obtaining (28), but this leaves only two equations with three unknowns, γ , T_0 and F .

At first sight it might seem that by taking two values of h and the corresponding two values of n from the experimental parabola, we could obtain two numerical equivalents of (28) which, together with (24), would allow solution for all three unknowns. This is a fallacy, however, because by dividing equations (18) and (22), the absolute value of N has been eliminated and the parabola now yields only two quantities.

It is therefore necessary to assume a numerical value for one of the three remaining unknowns, the one chosen being that which can best be assessed. The temperature T_0 is obviously the one to be chosen, as it is not likely to be outside the limits of 350°K. and 700°K. , i.e. it varies through no more than a factor of two. We shall not, however, assume a fixed value of T_0 , but apply the theory for a number of values of T_0 between these two limits, deferring discussion of its actual value until later.

Algebraic solution of equations (24) and (28) is impracticable as a routine method, and resort is therefore made to a graphical method, which can be applied quickly to most conditions met with in the F_2 region. It should be remembered that the final equations are only true if $dn/dt = 0$, a condition which is well fulfilled for the F_2 region below its maximum from 08h. to 16h. South African Standard Time, as previously pointed out.

3 (ii). Graphical Solution for F and γ

The value of n for substitution in equation (28) may be read at any height below h_M , where the electron density is greater than $0.3N$. These limits are those beyond which the curve differs appreciably from parabolic form, and between which all the virtual heights are originally scaled.

In order to construct a family of curves of $(n/N, \gamma)$ for various values of F from equation (28), the value of $(h-h_0)/T_0$ must be settled. A value of h agreeing with the conditions outlined in the previous paragraph is 280 km., as

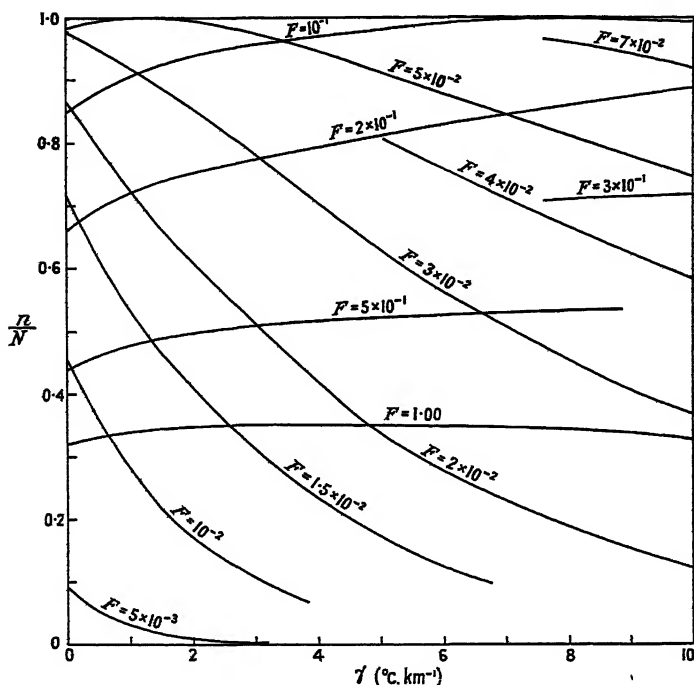


Figure 7. Curves of $\frac{n}{N}$ against γ for $\frac{h-h_0}{T_0} = \frac{1}{5}$, from equation (30).

may be verified by an examination of the electron density map. Assuming that T_0 will be of the order of 400°K ., we find

$$\frac{h-h_0}{T_0} = \frac{280-200}{400} = \frac{1}{5}. \quad \dots\dots(29)$$

When we recall that $C=16.2^\circ \text{C. km}^{-1}$, and substitute in (28), dropping the subscript b, this simplifies to

$$\left(\frac{n}{N}\right)^2 = \left(1 + \frac{\gamma}{5}\right)^{-(1+16.2/\gamma)} \left\{ F \left(1 + \frac{\gamma}{16.2}\right)^{-(1+\gamma/16.2)} \right. \\ \left. \times \exp \left\{ \left(1 + \frac{\gamma}{16.2}\right) - \frac{1}{F} \left(1 + \frac{\gamma}{5}\right)^{-16.2/\gamma} \right\} \right\}. \quad \dots\dots(30)$$

Choosing values of γ which give simple numerical values to the indices ($\gamma=1.62, 3.24, 4.86, 6.49, 8.10, 9.7$), and using equations (15) and (21) for $\gamma=0$, various values of F may be substituted to give n/N greater than 0.3. In this way the family of curves shown in Figure 7 was drawn.

Equation (24) may be written, with the usual value of C ,

$$\frac{h_M - h_0}{T_0} = \frac{1}{\gamma} \left[\left\{ F \left(1 + \frac{\gamma}{16.2} \right) \right\}^{-\gamma/16.2} - 1 \right]. \quad \dots\dots (31)$$

With the values of γ used before a similar family of curves may be drawn, and is shown in Figure 8.

These curves exhibit a number of interesting points:

- (i) The occurrence of maxima in the curves of n/N for $F = 5 \times 10^{-2}$ and $F = 10^{-1}$ for values of γ equal to 1.2 and 8 deg.km⁻¹ respectively indicate that under these conditions n becomes equal to N , i.e. the maximum electron density passes through the height h for which $(h - 200)/T_0 = \frac{1}{5}$.

As was illustrated by Figure 5, an increase in F corresponds to a lowering of the height, and the curves for values of F greater than 10^{-1} refer entirely to values of n above h_M , and therefore cannot be used for practical purposes.

- (ii) In Figure 8 the curves below $(h_M - 200)/T_0 = 0.2$ again refer to conditions when the maximum is below the selected height h , and therefore have no practical value.

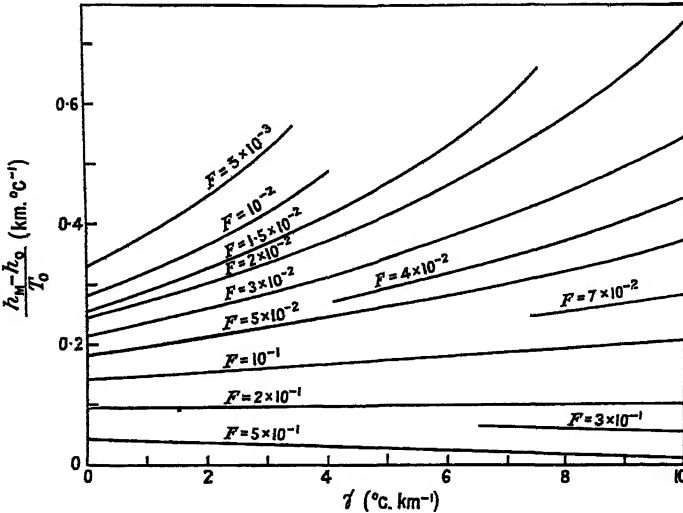


Figure 8. Curves of $\frac{h_M - h_0}{T_0}$ against γ for $\frac{h - h_0}{T_0} = \frac{1}{5}$, from equation (31).

The simultaneous graphical solution of equations (30) and (31) may now be accomplished, if values taken both from Figure 7 and from Figure 8 are plotted on the same diagram, using as coordinates the variables F and γ or, preferably, $\log F$ and γ . This has been done in Figure 9, where lines of constant n/N from Figure 7 are drawn in full and those of constant $(h_M - 200)/T_0$ from Figure 8 are shown broken.

The process of evaluating γ and F for a given experimental parabola now resolves itself into the following series of simple operations:

- (i) select a reasonable value of T_0 ;
- (ii) estimate the value of h by substituting this value of T_0 in the equation $(h-200)/T_0 = \frac{1}{5}$;
- (iii) from the parabola, find n at this value of h , and also N ;
- (iv) hence calculate n/N ;
- (v) from the parabola read off h_M ;
- (vi) evaluate $(h_M-200)/T_0$, using the value of T_0 chosen in (i);
- (vii) on Figure 9 locate the point which satisfies simultaneously (iv) and (vi), and read off the values of γ and $\log F$.

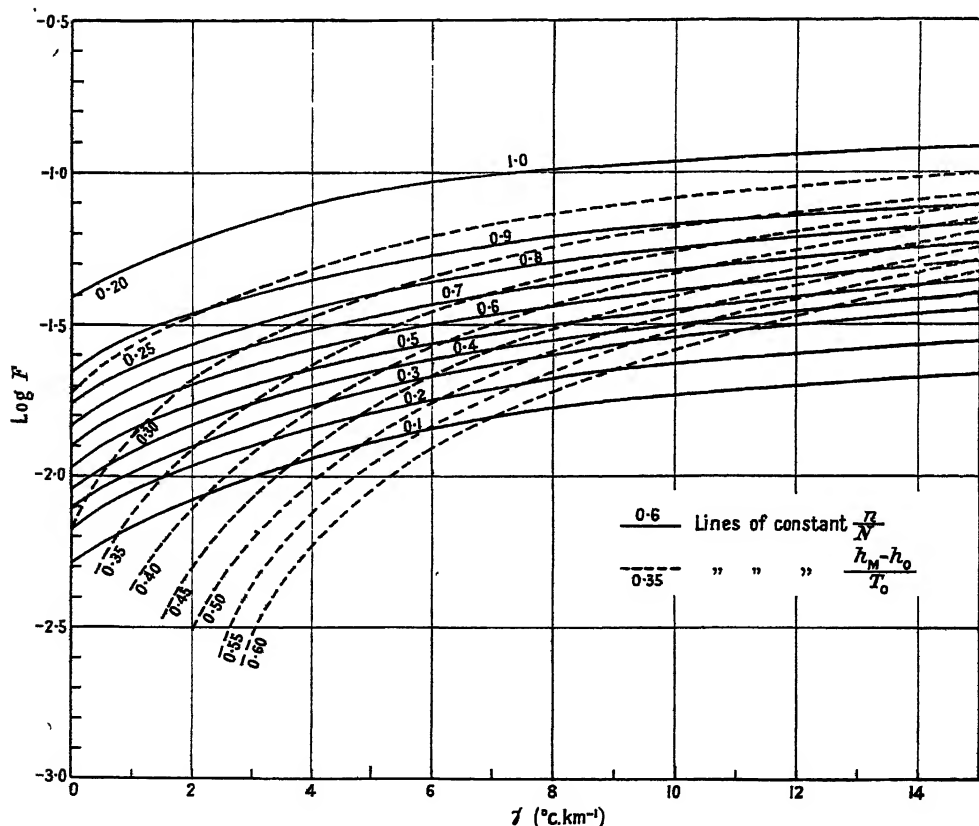


Figure 9. Graph for the evaluation of γ and F from experimental data.

The whole procedure is repeated for each value of T_0 selected.

The curve $n/N=1.0$ represents the maximum electron density, and thus the part of the graph above this is not shown, as it has no use. It coincides with the line $(h_M-200)/T_0=0.2$, which also represents the electron density maximum.

The angle between the two families of curves increases as we estimate n further below the maximum, and therefore values of γ and F estimated from values of n scaled near the maximum are not very accurate. This could be

avoided by choosing the fraction $(h-200)/T_0 \neq \frac{1}{5}$, but this would introduce other difficulties by putting n below $0.3N$ in the normal applications.

The two families of curves tend to run parallel for values of γ greater than 15 deg.km^{-1} , and this introduces uncertainty in values estimated in this region. Since for $\gamma = 15 \text{ deg.km}^{-1}$ and $T_0 = 250^\circ \text{K.}$ (a very low estimate), the temperature at 350 km. would be $2,500^\circ \text{K.}$, which is more than anyone has hitherto dared to suggest, this is no disadvantage. Godfrey and Price (1937) have shown by considerations of radiation equilibrium that this temperature cannot rise above $3,300^\circ \text{K.}$

3 (iii). *Treatment of Experimental Results*

The process just developed was applied to the experimental results for November 1945–February 1946, described in § 2 of this paper. The value of T_0 being unknown, the estimation of γ and F was repeated for each parabola for a series of values of T_0 , varying in steps of 50°K. from 250 to 600°K. In many cases the higher values of T_0 gave values of γ well off the graph, and these, in accordance with the previous discussion, have not been taken into account.

To illustrate the results of this treatment and the following discussion, Table 2 is given. It shows, in the first two rows of each section, the figures for γ and F for November. Similar tables were prepared for the other three months.

The values of γ obtained are practically all above 5 deg.km^{-1} which points straight away to temperatures of the order of $1,000^\circ \text{K.}$ in the F_2 layer. Further discussion on this point is, however, premature, because we have no means of selecting any particular value of T_0 as yet.

Some indication of the correct value of T_0 and γ may be obtained by performing two tests.

Test I.

From equation (17) we find

$$\beta = \frac{C \cos \chi}{n'_0 T_0 F}. \quad \dots\dots (32)$$

As the maintenance of the layer is presumably due to ionization of a particular species of atom or molecule, β should be constant. $\cos \chi$ may be calculated from the Nautical Almanac and we have a value of F for each T_0 . However, n'_0 will also vary with T_0 . During the preliminary work this was calculated from a probable value at 100 km. on the assumptions that there the temperature remained unchanged at 360°K. throughout the day and that there was a linear temperature gradient between the 100 and 200 km. levels. Later a more detailed examination of the problem gave what is believed to be a better value for the 200 km. level, $n'_0 = 1.5 \times 10^{11} \text{ cm}^{-3}$ when $T_0 = 360^\circ \text{K.}$, but did not indicate the variation of this quantity with T_0 . For the present purposes this quantity has been taken as a datum and modified for other values of T_0 by assuming it to vary in the same proportion as the earlier estimate. The values so obtained are shown in Table 3.

In this way the values of β corresponding to each T_0 at each hour were evaluated, and those for November are given in the third row of each section of Table 2.

Table 2. Experimental values of γ and F and derived values of β and B/n'_0 .
November 1945

S.A.S. Time	T_0 ($^{\circ}$ K.)	250	300	350	400	450	500	550	600
8 h. $\cos \chi = 0.529$	γ	11.7	10.7	9.2	8.5	8.0	9.5	11	—
	$100F$	3.94	4.34	4.68	5.25	5.85	7.50	9.33	—
	$\beta \times 10^{17}$	10.6	6.04	3.63	2.41	1.82	1.17	0.78	—
	B/n'_0	4940	2740	1440	901	675	529	396	—
9 h. $\cos \chi = 0.701$	γ	11.4	9.9	8.9	7.9	7.9	6.5	10	—
	$100F$	3.71	4.02	4.49	4.87	5.70	5.89	8.71	—
	$\beta \times 10^{17}$	15.0	8.65	5.00	3.45	2.48	1.97	1.10	—
	B/n'_0	6030	3120	1660	1080	815	565	461	—
10 h. $\cos \chi = 0.836$	γ	13.5	12.7	11.6	10.7	9.2	11.5	13.0	—
	$100F$	4.36	5.01	5.46	5.92	6.03	7.76	9.12	—
	$\beta \times 10^{17}$	18.1	8.27	4.88	3.39	2.80	1.78	1.27	—
	B/n'_0	8170	4240	2370	1550	1160	942	773	—
11 h. $\cos \chi = 0.927$	γ	11.3	9.65	8.3	7.6	7.1	7.0	5.5	—
	$100F$	3.21	3.36	3.61	4.01	4.52	5.21	5.25	—
	$\beta \times 10^{17}$	22.8	13.7	8.25	5.56	4.12	2.94	7.73	—
	B/n'_0	13100	6790	3550	2290	1690	1220	890	—
12 h. $\cos \chi = 0.966$	γ	10.2	9.3	8.0	7.2	6.6	6.7	5.5	9.8
	$100F$	2.69	3.14	3.36	3.76	4.20	4.98	5.07	8.32
	$\beta \times 10^{17}$	28.3	15.3	9.20	6.18	4.65	3.21	2.62	1.37
	B/n'_0	17200	8470	4520	2860	2110	1510	1110	920
13 h. $\cos \chi = 0.952$	γ	8.5	8.2	7.1	6.0	6.0	5.5	4.0	9
	$100F$	2.11	2.67	2.92	3.19	3.82	4.22	4.17	7.94
	$\beta \times 10^{17}$	35.8	17.6	10.4	7.17	5.01	3.74	3.12	1.40
	B/n'_0	15600	8070	4210	2600	1920	1390	993	812
14 h. $\cos \chi = 0.884$	γ	9.8	8.9	7.5	6.8	6.1	5.4	3.6	8.6
	$100F$	2.69	3.10	3.29	3.73	4.13	4.44	4.09	8.03
	$\beta \times 10^{17}$	25.6	14.1	8.56	5.70	4.31	3.29	2.96	1.29
	B/n'_0	13300	6990	3680	2320	1690	1210	895	720
15 h. $\cos \chi = 0.769$	γ	7.8	7.4	6.3	5.6	4.8	3.8	8	—
	$100F$	2.25	2.79	3.09	3.55	3.86	4.05	7.59	—
	$\beta \times 10^{17}$	27.0	13.6	7.96	5.19	4.01	3.13	1.39	—
	B/n'_0	11200	5770	3000	1900	1370	960	726	—
16 h. $\cos \chi = 0.613$	γ	6.5	6.1	5.2	4.8	3.6	5	12	—
	$100F$	2.01	2.57	2.98	3.52	3.69	5.62	10	—
	$\beta \times 10^{17}$	24.0	11.8	6.59	4.07	3.35	1.81	0.84	—
	B/n'_0	9020	4520	2310	1440	1040	718	689	—

 Table 3. Values of n'_0 for various T_0

T_0 ($^{\circ}$ K.)	250	300	350	400	450	500	550	600
$n'_0 \times 10^{-11} \text{ cm}^{-3}$	0.935	1.20	1.45	1.69	1.90	2.07	2.27	2.44

Test II.

The second test depends on the evaluation of the constant B , previously eliminated by division of equations (18) and (22). B is defined by equation (16) and is also given by a rearrangement of equation (22). From its definition it is seen that B is not a constant, since it contains n'_0 which is a function of T_0 . The quantity

$$\begin{aligned} \frac{B}{n'_0} &= \frac{\beta I_\infty}{\alpha w} \text{ from equation (16)} \\ &= \frac{N^2 \exp(1 + \gamma/C)}{n'_0 \left\{ F \left(1 + \frac{\gamma}{C} \right) \right\}^{(1+\gamma/C)}} \text{ from equation (22)} \quad \dots\dots (33) \end{aligned}$$

should however be constant at all times and for all months. The values for November are shown in the fourth row of each section of Table 2.

3 (iv). *Discussion of the Tests*

The correct value of β is very doubtful. Table 2 and the corresponding tables for the other three months have been drawn up to cover temperatures T_0 from 250°K. to 600°K. , which range includes most of the rough estimates found in the literature, and, if this range is accepted as the most probable for T_0 , the correct value of β should appear in the third row of each section. An examination of the tables shows that values of β in the neighbourhood of $3 \times 10^{-17} \text{ cm}^2$ fulfil this condition. Table 4 shows the values of T_0 , interpolated to the nearest 10°K. from the tables, which correspond to $\beta = 3.2 \times 10^{-17} \text{ cm}^2$.

Table 4. Values of T_0 ($^\circ \text{K.}$) corresponding to $\beta = 3.2 \times 10^{-17}$

S.A.S. Time (hours)	November	December	January	February
8	370	320	320	320
9	410	380	350	360
10	410	410	470	390
11	480	540	410	430
12	500	490	470	520
13	540	460	430	450
14	500	450	420	440
15	490	430	390	410
16	450	390	330	390

In theory, if a value of B/n'_0 is chosen to give the same T_0 as the value $\beta = 3.2 \times 10^{-17} \text{ cm}^2$ at a certain time, this same value of B/n'_0 should correspond to the identical set of values of T_0 shown in Table 4. However, B contains α and I_∞ , which may vary with season, sunspot activity, etc., respectively. Hence we shall not expect exact agreement in practice. To find the value of B/n'_0 which best corresponds to $\beta = 3.2 \times 10^{-17} \text{ cm}^2$, a figure was estimated to correspond to each of the noon values of T_0 in Table 4, and the mean of the four was taken. This gave the best value of B/n'_0 as 910 cm^{-3} ; the agreement was not good, the largest deviation from the mean being about 50%. Considering the relatively large percentage change in B/n'_0 corresponding to a change of only 50°K. in T_0 , however, this may be regarded as satisfactory. The T_0 values

corresponding to $B/n'_0 = 910 \text{ cm}^{-3}$, interpolated for the last row of Table 2 and the corresponding ones for the other three months, are shown in Table 5.

Table 5. Values of T_0 ($^{\circ}\text{K.}$) corresponding to $B/n'_0 = 910 \text{ cm}^{-3}$

S.A.S. Time (hours)	November	December	January	February
8	400	330	300	320
9	430	330	330	400
10	510	350	350	450
11	550	440	350	450
12	600	410	370	550
13	570	410	350	560
14	550	420	350	550
15	510	390	340	550
16	470	380	310	460

The agreement between the two tests is very pleasing, as they rest on independent grounds; therefore it does not seem out of place to take the mean values from these two tables the best representative values for further discussion. These are given in Table 6.

Table 6. Mean values of T_0 ($^{\circ}\text{K.}$)

S.A.S. Time (hours)	November	December	January	February
8	385	325	310	320
9	420	355	340	380
10	460	380	410	420
11	515	490	380	465
12	550	450	420	535
13	555	435	390	505
14	525	435	385	495
15	500	410	365	480
16	460	380	320	425

The absolute values of these T_0 's depend, of course, on the arbitrary choice of β as $3.2 \times 10^{-17} \text{ cm}^2$. The effect of changing this was investigated by estimating new values of T_0 by the two tests with $\beta = 1.6 \times 10^{-17} \text{ cm}^2$ and $8.0 \times 10^{-17} \text{ cm}^2$ respectively, and the corresponding figures for B/n'_0 . This resulted in an increase of the estimated T_0 by about 100°K. in the first case and a decrease by the same amount in the second case, but it did not affect the form of the variation of T_0 with time of day, as may be easily verified by an examination of Table 2.

Previous estimates have been made of the mean temperature at 200 km., but they vary widely in order of magnitude. Martyn and Pulley (1936), from considerations of electron collision frequencies and of auroral spectra, have tentatively adopted the value of $1,000^{\circ}\text{K.}$ at 200 km. in the daytime; Hulburt (1939b) has assumed the value 360°K. above 100 km. for lack of better evidence; while Vassy and Vassy (1942), assuming complete dissociation of both oxygen and nitrogen, obtained a temperature gradient which gives T_0 at 200 km. as 600°K. in summer. The values in Table 6, having a mean of about 430°K. , are well in accord with these estimates.

It may be seen that T_0 rises towards the middle of the day and falls off with some delay in the afternoon. This is true no matter what β is chosen to be.

It shows that the heating of the atmosphere does take place even at a height of 200 km., the temperature rising by about 30% of its value at 8 hours.

Estimates have been made of the 'temperature of the F_2 region' during summer daytime:

- (i) Appleton and Naismith (1935), from considerations of the variation in N between summer and winter noon, estimate that the temperature cannot be less than $1,200^\circ \text{K}$. in summer.
- (ii) Fuchs (1936) has concluded from estimated scale heights that the temperature of the F_2 region may rise as high as $1,900^\circ \text{K}$.
- (iii) Martyn and Pulley (1936) confirm Appleton's result of a temperature of $1,200^\circ \text{K}$.
- (iv) Vassy and Vassy (1942) assuming complete dissociation of O_2 and N_2 , obtain a summer temperature of $1,080^\circ \text{K}$. from the estimated scale height of 70 km. given by Appleton.
- (v) Godfrey and Price (1937), by a consideration of the equilibrium between absorption and secondary radiation, conclude that the temperature is of the order of $1,000^\circ \text{K}$., and cannot exceed $3,300^\circ \text{K}$.

For comparison, the temperature at 300 km. was calculated from the T_0 's of Table 6 and the corresponding γ 's. The individual variations of these figures for T_{300} are probably not significant of actual conditions, and therefore only the means are included in Table 7.

Table 7. Mean Temperatures at 200 and 300 km.

Month	November	December	January	February
Mean T_0 ($^\circ \text{K}$.)	485	408	369	447
Mean T_{300} ($^\circ \text{K}$.)	1097	1376	1392	1305

It will be observed that the mean values for T_0 are smaller for the midsummer months December and January than for November and February, while the reverse holds for temperatures at 300 km. This gives valuable support to the expansion theory, for, near the solstice, when heating is greatest, the atmosphere at 200 km. becomes heated in the middle part of the day, and is forced upwards by expansion, leaving cooler gas to take its place. The heated mass of gas would produce greater absorption of the solar radiation, thus preventing it from reaching the 200 km. level, where it would normally cause more heating.

As a last interesting point, we may make an estimate of the intensity of active incident radiation I_∞ above the earth's atmosphere. From equation (16)

$$I_\infty = \left(\frac{B}{n_0'} \right) \frac{\alpha w}{\beta}. \quad \dots (34)$$

We have assigned to B/n_0' the value 910 cm^{-3} from the results of test I. The value of α is open to discussion, and seems to vary with ion density and height, according to Wells and Shapley (1946); a reasonable mean seems to be $5 \times 10^{-10} \text{ cm}^3 \text{ sec}^{-1}$; β has been fixed at $3.2 \times 10^{-17} \text{ cm}^2$; and the value of the ionization potential w is given by Bacher and Goudsmit (1932) as 13.55 ev. for atomic oxygen and 14.48 ev. for atomic nitrogen. As a mean 14 ev. has been used. Thus we find

$$I_\infty = 0.313 \text{ erg.cm}^{-2} \text{ sec}^{-1}.$$

Hulburt (1939b) has calculated that the total black body radiation of energy per quantum greater than 13.7 ev. received from the sun is 8×10^{-3} erg.cm⁻²sec⁻¹, which is much lower than the above estimate.

Substituting back the value 8×10^{-3} in equation (34), we find $B/n'_0 = 23.2$, and from Table 2 this will be seen to correspond to T_0 at least as high as 2,000° K. This is ridiculous, and we can only conclude that the sun does not behave as a black body in this spectral region. A similar conclusion has been drawn by other workers (Bradbury 1938, Saha 1937, McNish 1937, Hunter 1943).

§ 4. CONCLUSIONS

The introduction of a linear temperature gradient into the theory of layer formation yields equations which may be applied with success for data for the F₂ layer, presented in the form of parabolae representing the distribution of ion density with height. The values of the temperatures at 200 and 300 km. obtained are in accordance with previous estimates, and are of a more quantitative nature.

The main weakness seems to lie in the assumption regarding the absorption coefficient β . In the first place its value is not fixed by the theory, and in the second place it varies with the frequency of the radiation absorbed, and an assumption of a mean value is not necessarily justified. A general treatment of this problem has been given by Chapman (1939). Again, the manner in which the recombination coefficient α varies with the temperature and pressure is not well established. Finally the theory is limited in its application by the assumption that $dn/dt = 0$ and by its neglect of the earth's curvature. Re-examination of these points seems to be the next step in its development.

It should be pointed out that the assumption of complete dissociation of oxygen and nitrogen above 100 km. is not a weakness, since a numerical recalculation is all that is required to apply the theory to an atmosphere containing constituents of known molecular weight.

ACKNOWLEDGMENTS

In conclusion it is a pleasure to thank Professor Varder for his constant encouragement and for making available the resources of his department, and the National Research Board and Council for a grant.

REFERENCES

- APPLETON, E. V., and NAISMITH, R., 1935, *Proc. Roy. Soc. A*, **150**, 685.
 BACHER, R. F., and GOUDSMIT, S., 1932, *Atomic Energy States* (New York : McGraw-Hill).
 BERKNER, L. V., WELLS, H. W., and SEATON, S. L., 1936, *Trans. Edin. Meeting Assn. Terr. Magn. Elect.*, p. 340.
 BOOKER, H. G., and SEATON, S. L., 1940, *Phys. Rev.*, **57**, 87.
 BRADBURY, N. E., 1938, *Terr. Magn. Atmos. Elect.*, **43**, 55.
 CHAPMAN, S., 1931, *Proc. Phys. Soc.*, **43**, 26, 483; 1939, *Ibid.*, **51**, 93.
 FUCHS, J., 1936, *Beitr. Geophys.*, **47**, 1.
 GODFREY, G. H., and PRICE, W. L., 1937, *Proc. Roy. Soc. A*, **163**, 228.
 HULBURT, E. O., 1928, *Phys. Rev.*, **31**, 1018; 1938, *Ibid.*, **53**, 344; 1939 a, *Ibid.*, **55**, 639; 1939 b, *Physics of the Earth*, edited by J. A. Fleming (New York : McGraw-Hill), Chap. X, p. 492.
 HUNTER, A., 1943, *Rep. Prog. Phys.*, **9**, 5 (London : Physical Society).
 MCNISH, A. G., 1937, *J. Appl. Phys.*, **8**, 718.
 MARTYN, D. F., and PULLEY, O. O., 1936, *Proc. Roy. Soc. A*, **154**, 455.
 MILLINGTON, G., 1932, *Proc. Phys. Soc.*, **44**, 580; 1935, *Ibid.*, **47**, 263.
 PIERCE, J. A., 1947, *Phys. Rev.*, **71**, 698.
 SAHA, M. N., 1937, *Proc. Roy. Soc. A*, **160**, 155.
 VASSY, A., and VASSY, M. E., 1942, *J. Phys. Radium*, ser. 8, **3**, 8.
 WELLS, H. W., and SHAPLEY, A. H., 1946, *Terr. Magn. Atmos. Elect.*, **51**, 401.

A New Method of Measurement of the Variation with Wavelength of the Refractive Index and Absolute Stress Optical Coefficients of Amorphous Solids

By J. E. H. BRAYBON
University College, Southampton

Communicated by S. Weintraub; MS. received 9th January 1950

ABSTRACT. White light is passed through a Michelson interferometer in one of whose arms a thin lamina of the substance in the shape of a tension member is supported by a loading frame. The emergent beam, examined by a spectrometer, shows a continuous spectrum crossed by interference fringes. For the unstressed lamina the position of the fringes gives the variation of refractive index with wavelength. For the stressed lamina the behaviour of the fringes gives the stress optical coefficients. No modification of the interferometer, and only three sets of observations on the fringes, are required. For the calculation of absolute values of the refractive index the values of two wavelengths must be known.

Full details of the method are given, and photographs of tests on Columbia Resin 39 are included. These show that CR39 becomes positively birefringent under tension.

§ 1. INTRODUCTION

THE method, which uses an unmodified Michelson interferometer, has been specially developed in order that measurements over a wide range of wavelength of both the refractive index and stress optical coefficients can be made on the same specimen. This is of importance in the case of a plastic, where the optical properties may vary with age and from sample to sample, and thus, unless the same specimen is used throughout, the refractive index cannot be correlated with the stress birefringence. The material used in checking the method was a condensed allyl ester, Columbia Resin 39 (CR39). In one arm of the interferometer, a thin lamina of the material (1 to 2 mm. thick), in the form of a tension member of uniform cross section, is held by a frame which may be used to apply a homogeneous tensile stress to the specimen. White light is passed through the system and the emergent beam is examined by a spectrometer. The continuous spectrum is observed to be crossed by a system of interference bands. From this system for the unstressed lamina, the variation of refractive index with wavelength is obtained; from the behaviour of the system when the lamina is stressed the stress optical coefficients are deduced. The theory of the method is given in detail.

§ 2. DETERMINATION OF REFRACTIVE INDEX DISPERSION CURVE

(i) *Experimental Details*

The material, in the form of a plane parallel thin lamina, is inserted in one arm of a Michelson interferometer. Light from a 'Pointolite' source P (Figure 1) is focused on an adjustable circular diaphragm I, collimated by the lens L and passes into the interferometer M. Emerging from M, it is focused as a vertical line by the cylindrical lens C on to the slit of a Hilger constant deviation spectrometer H, or a spectrograph.

If great care is taken to align all the components strictly along the optical axis of the system, then, for a suitable setting of the tilt of the interferometer mirror m_2 , the resultant spectrum is observed to be crossed by vertical dark interference fringes. The relation between fringe wavelength λ , refractive index μ , and thickness t , of the lamina may be expressed as

$$n\lambda = D - 2(\mu - 1)t, \quad \dots\dots(1)$$

where D is the optical path difference between the two interfering beams in the absence of the specimen and n , the fringe order, is an integer.

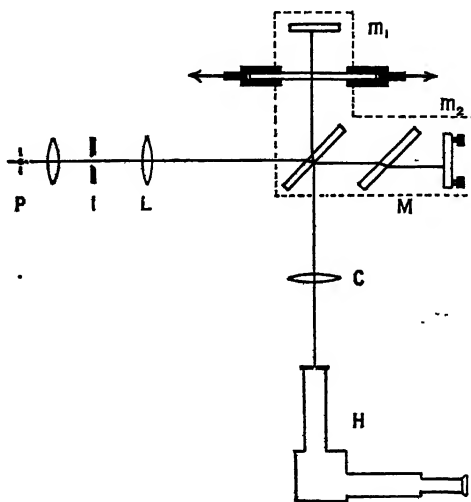


Figure 1. Optical system.

Due to the dispersion of μ with λ the fringes are not equidistant in wave number but are approximately symmetrical about a wavelength which is dependent upon D , i.e. upon the setting of the movable mirror, m_1 , of the interferometer. The photograph, Figure 2 (Plate I), obtained with a lamina of CR39, thickness 1.68 mm., shows the variation in position of this central fringe with alteration of D by lateral movement of m_1 . Further, replacing the lens C by a spherical lens, the diaphragm I by a vertical slit, and the source by a line filament, circular fringes are obtained and, in the region of the central fringe, it may be confirmed that corresponding fringes on either side of the central fringe are of the same order. Also, from equation (1), it is apparent that the fringe order rises to a maximum at the central fringe.

(ii) Calculation of the (μ, λ) Dispersion Curve

The absolute value of the fringe order cannot be determined directly without knowing the other variables in equation (1), but by plotting the fringe order (arbitrary) against λ , a curve of the form shown in Figure 3 can be obtained. From this curve the dispersion curve of μ against λ may be obtained as follows: Let the wavelengths of two fringes, on opposite sides of the maximum, but having the same fringe order, x (absolute value unknown), be λ_0 and λ_0' . If the wavelength of any fringe of order $(x + m)$ is λ_m , and μ_0, μ_0' and μ_m are the values of the refractive

index at λ_0, λ_0' and λ_m respectively, then, from equation (1), by elimination of x and D :

$$\mu_0 - \mu_m = (\mu_0 - \mu_0')(\lambda_m - \lambda_0)/(\lambda_0' - \lambda_0) + m\lambda_m/2t. \quad \dots\dots(2)$$

If μ_0 and μ_0' are known, μ_m is given by the above equation. However, it is not necessary in practice to calculate μ_m directly from equation (2), as the (μ, λ) dispersion curve can be obtained by a simple graphical construction.

In the absence of the $m\lambda_m/2t$ term, equation (2) represents the broken straight line of Figure 3 (b), through the points (λ_0, μ_0) , (λ_0', μ_0') , which is easily constructed. By adding to the ordinate of any point on this line the corresponding value of $m\lambda_m/2t$ the corresponding point on the dispersion curve is at once obtained.

It should be noted that this interpretation of equation (2) confirms that the fringe order rises to a maximum, since, in order to obtain the normal dispersion curve, the term $m\lambda_m/2t$ must be positive in the range λ_0 to λ_0' , and negative outside this range.

For this method, either μ_0 and μ_0' must be determined by other means, or D and one value of refractive index must be known. With the interferometer used

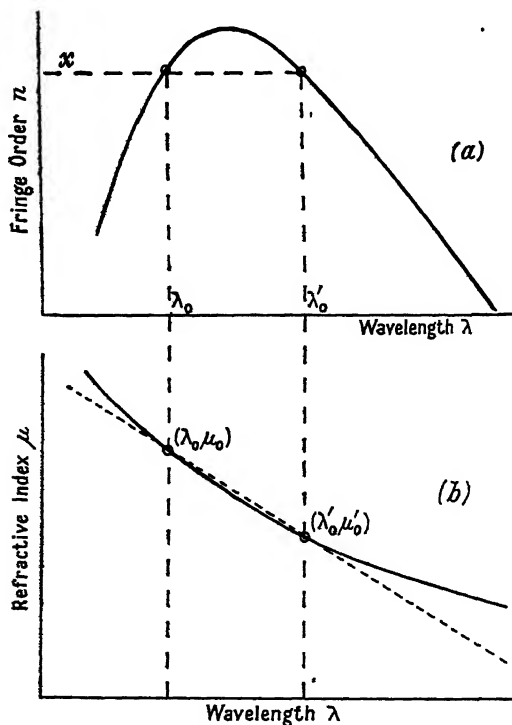


Figure 3. (a) Variation of fringe order (n) with wavelength λ .
(b) Variation of refractive index (μ) with wavelength λ .

it was more convenient to adopt the former course. Furthermore, unless the lamina itself is absolutely plane parallel, the interferometer, when the lamina is removed, will not give fringes without an alteration in the tilt setting of the mirrors in addition to a change in D . This completely precludes a determination of

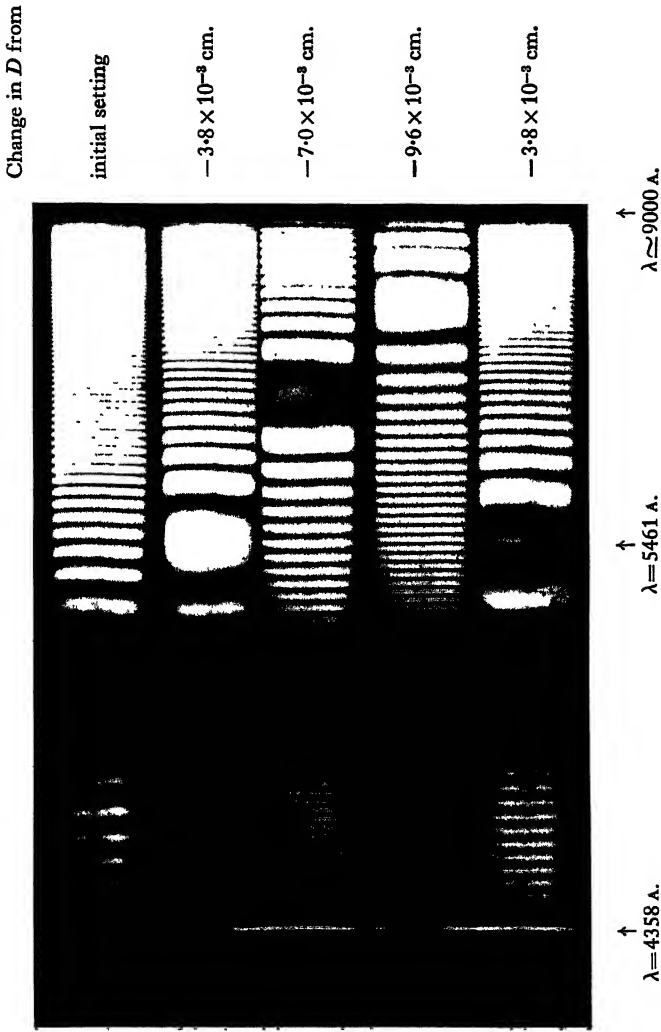


Figure 2. Photograph (positive print) of fringe systems obtained with various carriage settings of the Michelson interferometer. Tension member in position indicated in Figure 1.

PLATE I.

D since an adjustment of the mirror tilt in turn affects D . Because of this μ_0 and μ_0' were determined on two other specimens, both in the shape of 60° prisms. Experience had shown that in the case of CR39 the use of other samples is permissible as its optical properties are quite constant. Such a course, however, would not be suitable for a plastic such as Catalin, which is very susceptible to time effects, and the determination of μ must be made in this case directly by a refractometer or by a method involving the measurement of the Brewsterian angle (Pfund 1941). The final dispersion curve obtained agrees very closely with readings made using the 60° prisms at various other points over the visible range. The lamina used was supplied by the makers from their normal range, and, apart from cutting it into the form of a link for the purpose of stressing it as described below, it required no treatment.

§ 3. MEASUREMENT OF THE STRESS OPTICAL COEFFICIENTS

(i) *Experimental Details*

The lamina used must be in the form of a tension member which is subjected to a homogeneous tensile stress by means of a simple bell crank loading device. On loading as indicated in Figure 1 the fringe system described above is displaced outwards from the central fringe and the fringes over a wide range of wavelength disappear as the load increases. However, fringes are again visible if a polarizing prism is placed in the beam of light entering M, and the prism turned so that the longitudinal axis of the specimen (and therefore the direction of the stress) is parallel or perpendicular to the direction of the electric vector.

These effects occur because the specimen becomes optically anisotropic under load with its axis of symmetry along the line of stress.

The birefringence induced by the stress results in the fringes observed in light polarized with its electric vector at right angles to the tension being displaced by a different amount from those observed in light polarized with its electric vector parallel to the tension. For a suitable stress, over a comparatively large wavelength range the minima of the fringes obtained in the one position of the polarizer correspond to the maxima obtained in the other position. Thus, in the absence of the polarizer, no fringes are visible. If the load is further increased the fringes again become visible, indicating that one system has continued to move relative to the other.

This is illustrated in the photograph, Figure 4 (Plate II), which was obtained by placing a Rochon double image polarizing prism in the collimated beam entering the interferometer, and a spherical lens in place of C for focusing the resultant two beams from M, polarized at right angles to each other, upon the spectrometer slit as two separate point images. A double fringe system is then produced. In the absence of any stress the two sets of fringes are in alignment. The fringes move away from the central maximum and move relatively to each other as the load is applied.

It is of interest to note that the spectrogram taken one minute after the removal of the load shows that the fringe systems have not yet returned completely to coincidence. This indicates some residual birefringence due to creep (Braybon 1949). That there was also a very small amount of birefringence present before the load was applied may be seen from close examination of the spectrogram taken prior to loading. This was due to the immediate past history of the specimen.

(ii) *Calculation of Stress Optical Coefficients*

Photoelastic properties of an amorphous solid are characterized by two stress optical coefficients B_1 and B_2 , which at any given wavelength may be defined (Mueller 1935) by

$$\left. \begin{aligned} (\mu_x - \mu) &= \delta\mu_x = B_1 P_z, \\ (\mu_z - \mu) &= \delta\mu_z = B_2 P_z, \end{aligned} \right\} \dots\dots(3)$$

where P_z is the pressure in the z direction.

μ is the refractive index of the unstressed medium and μ_x and μ_z are the resultant indices of refraction in the stressed substance for light with its electric vector perpendicular and parallel to the direction of pressure. The sign of P_z is thus negative for a tension. By subtraction:

$$B = B_1 - B_2 = (\mu_x - \mu_z)/P_z, \dots\dots(4)$$

where B is the relative stress optical coefficient. Most determinations are of the stress birefringence and merely give the constant B .

At any given wavelength, the values of $\delta\mu_x$ and $\delta\mu_z$ may be deduced, for differentiation of equation (1) keeping λ constant gives

$$\delta\mu = -(\lambda/2t)\delta n. \dots\dots(5)$$

The changes in fringe order, $\delta n_x, \delta n_z$, for light polarized with its electric vector respectively perpendicular and parallel to the stress P_z , are found by comparing (fringe-order, λ) curves obtained as above with those obtained in the unstressed condition. The corresponding refractive index changes are therefore:

$$\left. \begin{aligned} \delta\mu_x &= -(\lambda/2t)\delta n_x = B_1 P_z, \\ \delta\mu_z &= -(\lambda/2t)\delta n_z = B_2 P_z, \end{aligned} \right\} \dots\dots(6)$$

and the resultant birefringence $\Delta\mu$ is given by

$$\Delta\mu = (\delta\mu_z - \delta\mu_x) = -(\delta n_z - \delta n_x)\lambda/2t = -BP_z. \dots\dots(7)$$

It should be noted that for a tension P_z is negative; for CR39 it was found that δn_x and δn_z are both positive and that $\delta n_x > \delta n_z$. Thus positive values of B_1 , B_2 and B are obtained, indicating that, as with the majority of other substances showing stress birefringent properties, CR39 becomes positively birefringent under tension.

Furthermore, if the total load applied to the specimen is L , equations (6) and (7) may be rewritten, with $1/2tP_z$ replaced by $b/2L$, where b is the breadth of the specimen, a quantity which may be determined with less error than t , which is of the order of only a millimetre.

The application of a load will result in a change in the dimensions of the specimen, i.e. in b and t . For the low values of load employed, however, it was considered unnecessary to correct for this variation which for the load used is of the order of 0.05%.

§ 4. DISCUSSION

The method for the determination of stress optical coefficients bears some resemblance to that of Mach (1872) and of Twyman and Perry (1922). However, in the first of these two methods a Jamin interferometer was used and the method depended upon the use of a compensator block of the same material and upon a

second measurement with specimen and compensator immersed in a fluid of known refractive index. In the second method a Michelson interferometer was used but the specimen was subjected to flexure and examined with monochromatic light.

The advantages of the present method are that only three sets of readings are required; measurements are made for small values of stress, reducing creep effects to a minimum; the specimen need not be very thin and fragile and its dimensions can be determined accurately. Estimation of the fringe shift for values of $\lambda \ll \lambda_0$ is somewhat difficult owing to the rapid rate of decrease of fringe order with decreasing wavelength, and results for the range may be somewhat less accurate. Obviously the method requires that the straight line through the points (λ_0, μ_0) and (λ'_0, μ'_0) be accurately determined and carefully drawn.

The method has been used successfully on CR39 over a wide range of wavelength. The work is being extended to other plastics and full details of the results will be published later.

ACKNOWLEDGMENTS

The author wishes to thank Professor A. M. Taylor for his advice and encouragement, and Mr. S. Weintroub for his assistance in the preparation of this paper.

REFERENCES

- BRAYBON, J. E. H., 1949, *Nature Lond.*, **163**, 689.
 MACH, E., 1872, *Ann. Phys., Lpz.*, Ser. II., **146**, 316.
 MUELLER, H., 1935, *Physics*, **6**, 179.
 PFUND, A. H., 1941, *J. Opt. Soc. Amer.*, **31**, 679.
 TWYMAN, F. and PERRY, J. W., 1922, *Proc. Phys. Soc. Lond.*, **34**, 151.

Vibrations of Free Elliptical Plates

BY MARY D. WALLER*

MS. received 16th January 1950

ABSTRACT. The systematic observations which have been made on the normal vibrations of free elliptical plates confirm Chladni's results. The nodal systems may be divided into the same four classes of symmetry as those of rectangular plates.

The correspondence which exists between the vibrations of elliptical, rectangular and circular plates is traced. Combined modes of vibration do not occur on elliptical plates as they occasionally do on rectangular plates.

§ 1. INTRODUCTION

THE only extensive experimental study of vibrating elliptical plates would appear to be that of Chladni. The 23 drawings given in *Die Akustik*, 1802, concern a plate whose axes are in the ratio of about 3/2 and the nodal systems are arranged systematically, as in the photographic records of the present paper. The text, §§ 142–157 includes data regarding the vibration frequencies of many other plates which may be converted from the notation of the chromatic scale by using the Zamminer table given in Rayleigh's *Theory of Sound*, I, 1894.

* Private address: 5 Gloucester Gate, Regent's Park, London N.W.1

Pavlik (1937) has used the Rayleigh-Ritz approximate method (Rayleigh 1894, 1911, Ritz 1909) to calculate the first seven natural frequencies of a rather wide stainless steel plate (ratio of axes 1.29) and to compare these with observation. The agreement is good for the lower frequencies, but for the sixth and last pair of frequencies the values given by theory exceed the observed values by 44% and 52% respectively. The theory, which involves laborious calculations, has not been tested on narrower ellipses.

It is evident that our information regarding the vibrations of elliptical plates must depend principally upon experiment, and it is desirable that photographs of the nodal designs produced should be placed on record.

The present work forms part of a more general study of the vibrations of plates which has not as yet reached completion. Details regarding the experimental technique have been given in previous publications (Waller 1949, and references therein).

The subject is approached through considerations of symmetry which make it easy to arrive at a general understanding of vibrations of plates of any given geometrical shape.

§ 2. SYMMETRY AND CLASSIFICATION OF THE NORMAL NODAL SYSTEMS

The elliptical plate possesses two mirror lines of symmetry, the major and the minor axis, and the rotational symmetry is 180° . Lines of symmetry divide the nodal pattern so that one half is the mirror image of the other. The fact that the lines of symmetry may be nodal or antinodal may be appreciated by examining the first nodal system of the first two rows of Figure 1 or 2 (see Plate). In the 1|1 system of the second row the lines are nodal and the displacements of corresponding points of the surface on either side of them are equal and out of phase. In the 0|2 system of the first row the lines are antinodal, that is to say they pass through the points of greatest displacement, and the displacements of corresponding points on either side of them are equal but in phase.

It is evident that all possible nodal systems can be divided into four classes according as (i) the centre is antinodal, (ii) the minor axis is nodal, (iii) the major axis is nodal, (iv) both the axes are nodal.

These classes correspond with those of rectangular plates where the two lines of symmetry are the two medians (Waller 1949).

All the nodal designs can be brought into self-coincidence by a rotation of 180° about the axis passing through the centre of, and perpendicular to the plane of, the plate.

§ 3. THE NORMAL NODAL SYSTEMS

Photographic records of the nodal designs obtained on a narrow plate (ratio of axes 2|1) and a wide plate (ratio of axes 5|4) are shown in Figures 1 and 2 respectively.

The arrangement of the nodal systems is the same as Chladni's and it corresponds also with that of rectangular plates (Waller 1949).

Each column, as indicated by the number above it, contains systems with a constant number of lines running in the direction of the shorter axis; and each row, as indicated by the number to the left of it, contains systems with a constant number of lines running in the direction of the longer axis. A nodal ellipse counts as two in the latter case. It will be seen that two nodal classes occur alternately in the first row and again in the third row, and the other two classes

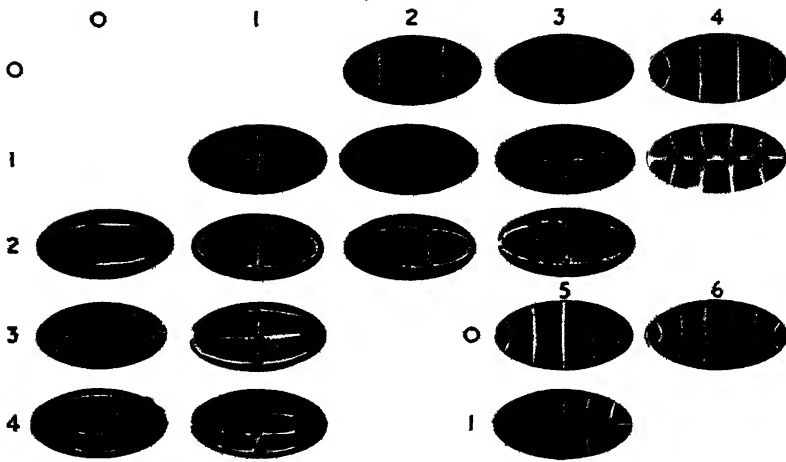


Figure 1. Ratio of axes $\frac{2}{1}$

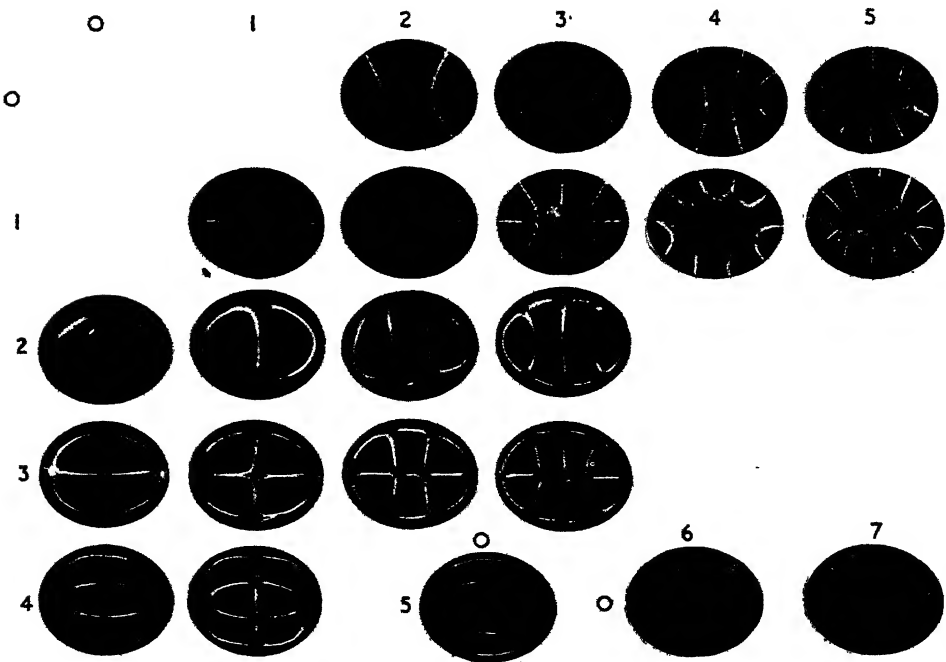


Figure 2. Ratio of axes $\frac{5}{4}$

Normal nodal systems of free elliptical plates.

in the second row and again in the fourth row. Three spaces at the top left-hand corner must remain vacant. The diagram can be continued indefinitely downwards and to the right.

§4. THE NATURAL FREQUENCIES

Table 1, which is in two parts, has been arranged so as to correspond with the nodal systems of Figures 1 and 2 respectively. The frequencies are expressed in terms of the gravest tone taken as unity. Particulars regarding the dimensions and actual lowest frequencies in plates used are also included. The numbers in italics are approximate, having been estimated by extrapolation from curves derived from the earlier frequencies of any given row or column. That the frequencies are in general agreement with Chladni's approximate results can be seen by comparing Table 1 with Table 2, which has been compiled from his data concerning plates of the same shape as those used in the present experiments.

Table 1. Relative Frequencies of Normal Vibrating Modes of Free Elliptical Plates

2/1 Ellipse (see Figure 1)						
	0	1	2	3	4	5
0			1*	2.58	4.7	7.3
1		1.77	3.27	5.68	8.29	11
2	4.25	6.57	9.43	12.6		
3	10.6	14				
4	17	22				

* Actual frequency of brass plate, major axis 12.68 cm., minor axis 6.39 cm., thickness 1.62 mm., was 438 c/s.

5/4 Ellipse (see Figure 2)							
	0	1	2	3	4	5	6
0			1*	2.45	4.28	6.66	9.39
1		1.07	2.59	4.34	6.8	9.6	13
2	2.03	3.99	6.71	10			
3	4.42	7.41	10.7	14			
4	9.01	12					
5	14						

* Actual frequency of brass plate, major axis 12.60 cm., minor axis 10.18 cm., thickness 1.62 mm., was 414 c/s.

Table 2. Chladni's Vibration Frequencies (approximate)*

2/1 Ellipse						5/4 Ellipse					
0	1	2	3	4	5	0	1	2	3	4	5
0		1	2.5	4.7	7.4			1	2.5	4.5	7
1	1.6	3.3	5.6	8.3	11		1.1	2.5	4.5	7	9.2
2	3.9	6.6	9.4	13	17	2	3.8	6.8	11	16	18
3	10	14	18	20		4.8	7.4	11	14	19	
4	19	24				8.9	12				
5						14					

* Converted from chromatic scale notation.

§ 5. FURTHER REMARKS

(i) *Comparison with Circular Plate*

On a circular plate, which is the limiting shape of elliptical plates, nodal ellipses become circular and nodal hyperbolic lines become diameters.

The correspondence between elliptical and circular nodal systems may be traced on Figure 2 by proceeding as follows:

Study two rows at a time, the first with the second and the third with the fourth. Couple each design in the upper of the two rows with the one preceding it in the lower of the two rows. Then each pair of nodal systems will be found to correspond to a single nodal system on the circular plate (Waller 1938).

For example, 0|2 in the first row and 1|1 in the second row both correspond to the two-nodal diameter system of the circular plate, 0|3 and 1|2 to three diameters, and so on.

Similarly pairs of nodal systems can be picked out in the third and fourth rows, omitting the nodal ellipse which occupies the first place and which corresponds to the nodal circle. Thus the 2|1 and 3|0 systems both correspond with a nodal circle with one diameter, and so on.

The ratios of the frequencies of paired nodal systems can be studied in Table 1, noting how much nearer they are to unity for the wider than for the narrower plate, 1.07 for example as compared with 1.77 for the first pair. In the limiting case of the circle the ratio becomes equal to unity.

(ii) *Comparison with Rectangular Plates and with Bar*

As previously noted, the nodal systems of elliptical and rectangular plates are arranged in the same way. The nodal designs evidently become more nearly alike as the plates become narrower. On very narrow plates the only nodal systems met with in practice are those which can be placed in the first row of Figure 1.

These approximate to the nodal systems of free bars where the lines run at right-angles to the length and are divisible into two classes according as the centre is antinodal or nodal.

(iii) *Combined Modes of Vibration*

We may ask why among the photographic records of Figure 2 the 1|4 and 2|3 designs are so distorted. Turning to the second part, Table 1, it is found that the corresponding vibration frequencies are nearly equal, 6.8 and 6.71, respectively. The fact that the two modes have attempted to combine, somewhat imperfectly, indicates that the plate is not quite uniform and explains why the designs are so irregular.

Incidentally this provides by far the most sensitive means of detecting a want of uniformity in a plate. But unfortunately it is unlikely to be of any practical use since combination occurs so rarely and cannot be predicted.

We will conclude by mentioning one important difference between the nodal designs of rectangular and elliptical plates.

On rectangular plates a trace of a second vibrating mode can sometimes be detected in a given nodal design, which always belongs to the same class of symmetry and which does not impair the symmetry of the design. Such combined vibrations have not been noticed on elliptical plates. This problem

of combined vibrations, due presumably to anticlastic coupling, is a difficult one. It can probably be studied most fruitfully by further experiments on plates of different shapes.

ACKNOWLEDGMENTS

I have again to thank the Council of the Royal Free Hospital School of Medicine and Dr. W. A. Leyshon for giving me the hospitality of the Physics Department and Mr. D. J. Morgan for kindly cutting the plates.

REFERENCES

- CHLADNI, E. F. F., 1802, *Die Akustik*, 2nd unaltered edition 1830 (Leipzig: Breitkopf u. Härtel), §§ 142-157, figs. 179-201.
 PAVLÍK, B., 1937, *Z. Phys.*, **107**, 458.
 RAYLEIGH, Lord, 1894, *Theory of Sound*, I, 2nd edition (London: Macmillan), Chapter X; 1911, *Phil. Mag.*, **22**, 225.
 RITZ, W., 1909, *Ann. Phys., Lpz.*, **28**, 737.
 WALLER, M. D., 1938, *Proc. Phys. Soc.*, **50**, 70; 1949, *Proc. Phys. Soc. B*, **62**, 277.

LETTERS TO THE EDITOR

Preparation of Alkali Metals in Glass

When studying the electrical conductivity of the alkali metals it is necessary, owing to their great chemical activity, to prepare specimens for experiment in sealed glass capillary moulds. It is usual to make these of common 'soft' glass so that platinum electrodes (which do not react with the metals or become oxidized during glass-blowing) may be sealed in. In recent work it has been necessary to use two special glasses for the preparation of potassium and lithium specimens respectively; we wish here to report our experiences with them.

The electrical resistance of 'pure' potassium specimens exhibited two distinct 'kinks' in the low temperature region between 20° and 4° K. (MacDonald and Mendelssohn 1948, 1950). To determine whether this was an intrinsic property of the ideally pure metal or due in some way to a small impurity content several distillations were carried out yielding purer samples. A limit, however, as judged by the residual resistance at very low temperatures, was reached quite soon. It was presumed that this was due to impurity (sodium probably) being collected from the glass distillation container; in order to minimize this, 'Pyrex' glass was used for making the container because of its lower sodium content. To effect further improvement a special glass (C 98) has been produced in which the only alkali ion present is potassium. This glass, of composition and properties shown in the accompanying Table, works well in a flame and can be drawn and blown easily. It is very

	SiO ₂	B ₂ O ₃	Al ₂ O ₃	BaO	Na ₂ O	K ₂ O	Li ₂ O	Expansion coefficient (0-400° C.)
C 98	64	2	—	18	—	16	—	9.6×10^{-6}
C 51 65	69.2	—	—	18.2	—	—	12.6	10.3×10^{-6}
C 10	2.3	26.8	29.6	35.0	4.4	1.9	—	9.0×10^{-6}

Components are given in percentages by weight.

resistant to devitrification (none appearing in five days at any temperature between 675° and 1,000° C.) and may be sealed directly to ordinary soda-lime-silicate glasses of about the same expansion coefficient. When this glass was used for the distillation vessel a useful improvement in residual resistance was immediately observed and, furthermore, the 'kinks' were now found to have disappeared, suggesting very strongly that the resistance phenomenon was due fundamentally to small amounts of sodium impurity, although the fundamental physical source of the phenomenon is still obscure.

When attempting to work with lithium metal a serious obstacle is met with since the molten metal reacts very strongly with ordinary glass, forming almost instantaneously a black compound on the surface of the glass; the action generally proceeds so rapidly as to cause fracture of the glass itself. It has been suggested (see, for example, Justi 1948) that the high melting temperature of lithium is primarily responsible for this behaviour; since, however, sodium can readily be heated in glass to temperatures well above this value without appreciable reaction, this hypothesis seems untenable.

It was thought at first that the process might be one of replacement of the sodium in the glass by lithium, and, therefore, a special glass was made up which contained lithium as the only alkali ion. The well-known Lindemann glass is of this type (containing only lithium, beryllium and boric oxides), but it is difficult to work into the apparatus required by the experiments. The simple lithium-barium-silicate glass (C 51/65) of composition shown in the Table can be drawn into tubing (including capillary) by a skilled glass-worker, and although the glass has a low softening temperature and must be worked with care, the tube can be sealed directly to ordinary soda-lime-silicate glasses of similar expansion coefficient. Thin platinum electrodes can be sealed into the glass without cracking.

Unfortunately, tests soon showed that this glass was no more resistant to lithium than ordinary 'soft soda' glass. Recalling, then, that fused silica is also attacked in the same way by lithium, it was thought that the silica in the glass was probably the responsible constituent. An experiment was therefore made using tubing of the type employed in the construction of sodium discharge lamps; the tubing has an inner lining of a borate glass which is almost free from silica, and is known to be resistant to the attack of both sodium and potassium vapours. The composition of the glass (C10) is also given in the Table. C10 has proved sufficiently resistant to permit lithium to be cast successfully in it if the metal is solidified fairly rapidly (say 10 to 20 seconds) after running into the mould; thereafter a blackening does begin to occur. It is hoped in the future to make experiments with other glasses in which the silica content will be reduced to zero and the Na_2O and K_2O replaced by Li_2O .

Clarendon Laboratory, Oxford.

D. K. C. MACDONALD.

British Thomson-Houston Co. Ltd.,
Research Laboratory, Rugby.
4th April 1950.

J. E. STANWORTH.

JUSTI, E., 1948, *Ann. Phys., Lpz.*, 3, 183.

MACDONALD, D. K. C., and MENDELSSOHN, K., 1948, *Nature, Lond.*, 161, 972; 1950, *Proc. Roy. Soc. A* (in the press).

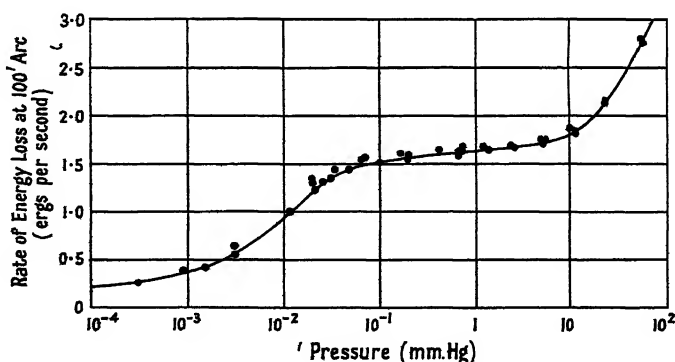
The Dissipation of Energy by a Pendulum Oscillating in Air at Low Pressures

The energy dissipation of a seconds pendulum has been measured by Atkinson (1938), who described experiments with different pendulums by means of which losses due to suspension, rod and bob were calculated. Such losses related to pendulums swinging in the open air, and these were found to be higher when the pendulums were inside cases. The loss of energy is due to four causes: (a) air resistance to the bob, (b) air resistance to the rod, (c) bending of the suspension spring, and (d) movement of the support. The combined loss due to causes (c) and (d) can be determined directly from measurements made with the pendulum in a vacuum.

The air pressure in the cases containing the 'Shortt clocks' is normally reduced to a value 30–20 mm. Hg, and it was decided to measure the losses of one of these pendulums at lower pressures and, by extrapolation to zero pressure, determine the losses due to the flexure of the suspension spring and the movement of the support.

The seconds pendulum of the standard Shortt clock seconds type, used for the tests, was carried by a spring ground from a solid strip of steel 2 mm. thick. The upper thick end of the strip was gripped between the two halves of a cylindrical trunnion while the lower thick end carried a steel pin which engaged with two hooks formed at the top end of the 8 mm. diameter invar pendulum rod. The bob, made of type metal, was a cylinder 9 cm. in diameter, 10 cm. long, weighing 6.5 kilograms. It was supported a little below its centre of gravity by means of an invar sleeve—pinned to the pendulum rod—carrying a brass collar of such a length that its upward expansion due to temperature compensated the corresponding downward expansion of the suspension spring and pendulum rod. The case was a copper tube of internal diameter 21.5 cm., fitted with heavy bronze flanges at the top and bottom carrying lugs by means of which it was solidly bolted to the wall. The trunnion, holding the top of the suspension, was supported by a four-legged bronze casting which stood on the inner edge of the top flange. This casting, which also carried the maintaining mechanism, was covered by a glass bell-jar, sealed to the outer edge of the flange. The base of the case was closed by a thick glass disc sealed to the bottom flange. When the pendulum was swinging, the underside of the cylindrical bob was 8.2 cm. from this plate, but the rod was extended downwards to carry a silver 'beat plate' which moved to and fro just clear of the glass. This beat plate was engraved with a series of cross-lines spaced 1.67 mm. apart which was the appropriate interval to represent 5' of arc. The movement of these lines, as the pendulum oscillated, was observed by means of a microscope. When illuminated by a suitably placed lamp, one of the plate lines came into view at the end of a swing, slowed to a standstill, and then swung 'out of focus' again, followed by a similar movement of another line entering from the opposite side. It was quite a simple matter to read off these 'standstill positions' on the eyepiece scale at any particular time or, alternatively, to observe the times when the lines 'stopped' at particular scale divisions, or at the same division.

The rate of energy loss of a swinging pendulum is $I\omega d\omega/dt$, where I is the moment of inertia of the pendulum and ω the angular velocity when passing through its equilibrium position. The value of I for the pendulum used was 6.852×10^7 g.c.s. units and, representing the total arc of swing by A , the rate of energy loss in ergs per sec. was $14.28A \, dA/dt$. To determine the relation between the rate of energy loss and the gas pressure it was only necessary to measure dA/dt at 100' total arc swing at various pressures.



The air pressure in the case having been adjusted to the required value, the arc of the pendulum was raised to 110' or more by utilizing the normal impelling mechanism; the impulsing was then cut off. The times were recorded at which the total arc became 105' and 95', respectively, and dA/dt determined. The gas pressure was measured by means of an oil manometer at the higher pressures and by a McLeod gauge at the lower values. The results obtained are shown in the accompanying diagram.

They indicate that the energy loss due to the flexure of the particular spring used, together with the losses due to the movement of the suspension brackets, i.e. of the head of the case, is less than 0.27 erg per sec., and probably reaches a limiting value of 0.20 erg per sec.,

that the damping at 3×10^{-4} mm. Hg is only about one-sixth the damping at 10^{-1} mm. Hg, and that gas viscosity damping still persists at a pressure of 10^{-4} mm. Hg, which agrees with the fact that the quartz fibre vacuum gauge is operative at this pressure. The limiting value of 0.20 erg per sec., which represents the suspension losses, is considerably less than that previously estimated by the author—a value criticized by Atkinson as being too high—but the earlier figure was based on tests which were confined to pressures above 1 mm. Hg. The present tests also confirm the statement by Loomis, quoted by Atkinson, that the decrement of a half-second pendulum was reduced to one-half by decreasing the gas pressure from 0.025 mm. to 0.001 mm. Hg.

Department of Physics,
University College, Exeter.
5th April 1950.

W. H. SHORTT.

ATKINSON, E. C., 1938, *Proc. Phys. Soc.*, 50, 721.

REVIEWS OF BOOKS

Higher Physics, by E. NIGHTINGALE. Pp xvi+808. (London: G. Bell and Sons Ltd., 1948.) 27s. 6d. Also in separate parts: Pt. I. *Mechanics and Properties of Matter*, 7s. 6d.; Pt. II. *Heat*, 7s. 6d.; Pt. III. *Light and Sound*, 10s.; Pt. IV. *Electricity*, 12s. 6d.

The clarity of style which has characterized Mr. Nightingale's many school texts in the past is in evidence in his latest effort. This book is of intermediate standard and is primarily intended for sixth form scholars.

The wide field of subject-matter which is covered within the confines of this single volume has necessitated considerable 'condensing' in certain sections, but on the whole this process has been carried out quite efficiently. However, it is inevitable that some subdivisions become unduly compressed when room is found for such topics as the theory of flight, the quantum theory of spectra, etc.

An undesirable effect of the lag, so prevalent in the post-war era, between the preparation of an MS. and its publication is exemplified in the present instance by the author's failure to record the recent official acceptance (October 1946) of the absolute system of electrical units as the practical system, to the exclusion of the International Scale. Also no mention is made, rather surprisingly, of the M.K.S. system of units.

The author makes frequent use of the methods of dimensions in the various sections of the book and the use of graphs in exhibiting experimental results is also stressed in the text.

The errors noted are remarkably few, but one which should be noted is the mention, on page 633, of the use of an iron former for the suspended coil of a galvanometer.

Line diagrams which are capable of reproduction by the student are freely employed throughout the text, and another feature which commends the book as one to be treasured by the scholarship candidate is the large selection of examination questions at the end of each chapter.

R. W. B. S.

Terrestrial Radio Waves: Theory of Propagation, by H. BREMMER. Pp. x+343. (New York, Amsterdam, London, Brussels: Elsevier Publishing Co., Inc.; British Agents: Cleaver-Hume Press, 1949).

The author asks us in his Preface to bear in mind that his aim is to describe the mathematical-physical methods for the computation of transmitter fields. The statement of the problem is presented briefly in the opening pages and looks deceptively simple: it is required to compute the radiation field of a Hertzian dipole above a spherical earth. The influence of the ionosphere and the refracting lower atmosphere is deferred until Part II of the book. It takes some hundred odd pages of concisely written mathematical analysis to arrive at the final formulae for the first part of the problem in a form suitable for numerical work, and the reader will readily agree with Nicholson that the problem was one of the most difficult at one time facing the theorist. The author and his colleague Professor van der Pol

were foremost in reducing its solution to a practical form for numerical work ; their original research papers, published in the *Philosophical Magazine* in 1937-1939, have long been recognized as the "textbook" of workers in this field. Indeed, much of Part I will be familiar to those who have read these papers.

The apparent ease with which the author systematically takes us through pages of complicated mathematical analysis could only come after long experience and familiarity with his subject; his emphasis on the physical interpretation at each stage is an outstanding example of the presentation of mathematical physics. Even so, it is to the experienced theoretician that his descriptive passages are directed, and it is unlikely that the average radio engineer has the necessary mathematical equipment for their appreciation. Nevertheless the debt owed by the radio engineer, as shown by his extensive use of the results of this work in the past, was aptly expressed by T. L. Eckersley in his concluding remarks to the Radio-location Convention in 1946 when he said that "there is nothing so practical as a good theory".

Without much preliminary, a beginning is made to the solution of the point source and spherical earth problem. Starting with the classical solution, in the form of a series of zonal harmonics, the author takes us through the various transformations and approximations with systematic thoroughness, eventually leading to the familiar residue series. An instructive account of the theories of steepest descent and stationary phase is of general interest in the relationship of ray to wave theory, and is more clearly discussed here than in most books on the subject. The author does well in giving prominence to the physical significance of geometric-optical approximations and of Watson's transformation, which might easily have remained as so much mathematics.

The concluding chapter of Part I will be particularly welcomed by the radio engineer whose interest is in the calculation of field strengths. The specific examples chosen and worked through in detail will be of assistance in using the formulae developed earlier in the book and now collected together in a convenient form for computation. Curves covering a wide range of frequencies are also included.

Part II of the book is devoted to a discussion of the influence of the ionosphere and refraction by the troposphere. With the simplifying assumption of a stratified atmosphere whose refractive index varies only with height, a comprehensive discussion is given of both the residue series and geometric-optical solutions. The reader will find an enlightening account in Chapter VII of the method by which the spherical earth problem can be replaced by one with a "flat" earth and a modified index of refraction. Closely allied to this is Eckersley's method of taking refraction into account to a first approximation by using a "fictitious" radius of earth which is usually larger than the actual radius. During the war years this theory has been supplemented by the more refined mode theory of super-refraction which explains the abnormally long ranges experienced by microwave radar during favourable meteorological conditions. The complexities of the eigenvalue problem involved in the evaluation of height functions receives detailed attention ; the discussion of the W.K.B. and the more accurate Krammer phase integral approximation is of general interest to the mathematical physicist.

This brief outline of the book cannot possibly do justice to its wealth of detail. It is without doubt a book for the propagation specialist who requires a fundamental understanding of his subject. There is, however, much of general interest for the theoretician.

W. WALKINSHAW.

Oscillations of the Earth's Atmosphere, by M. V. WILKES. Pp. ix+74. Cambridge Monographs on Physics. (Cambridge: University Press, 1949). 12s. 6d. net.

This excellent little volume by the Director of the University Mathematical Laboratory, Cambridge, deals with the oscillations produced daily in the atmosphere by the sun and the moon, as revealed mainly by the solar and lunar daily variations of barometric pressure. The subject may be said to have started with Torricelli's invention of the barometer and Newton's gravitational theory of the tides, greatly developed by Laplace, who in his last years devoted much attention to atmospheric oscillations. Kelvin showed the importance of resonance in greatly magnifying the semidiurnal atmospheric oscillations due to the sun's thermal and

tidal action, so that the lunar tide in the barometer, unlike that in the sea, is relatively inconspicuous. The present data on the subject are described, and also Laplace's tidal theory. The last and most novel part of the book is Chapter V which gives the modern theory of these oscillations and their resonance, a theory begun by Taylor and Peteris and since then much developed by the author and Dr. K. Weekes. The book should interest many physicists, mathematicians and meteorologists and stimulate further developments of which the subject still stands in much need.

S. C.

Sound (A University Text-book of Physics, Volume II), by J. H. POYNTING and J. J. THOMSON, revised by W. S. TUCKER. Pp. viii + 251. Tenth Edition. (London: Charles Griffin, 1949.) 20s.

During the last 50 years nine editions of this well-known textbook have been published. It is not surprising that in this present edition the reviser has found it necessary to write effectively a new book. The text has been rearranged and the theory simplified by the use of calculus methods. The reviser claims that the book has been written to remedy a deficiency in present University courses, where 'Sound' "is frequently relegated to the end, and is largely crowded out", and has been planned for students taking honours courses in branches other than physics, viz. those who have chosen physics as a subsidiary subject for their degrees.

The text mostly divides into two parts—pure and applied acoustics. The former follows normal procedure and requires little comment except that an improvement could be made in the definition of terms introduced and in the physical-mathematical analysis relation. In the applied acoustics part the question of relevancy of subject matter is much a question of opinion but it would appear that the text would gain in clarity and cohesion by considerable pruning and rearrangement. A considerable proportion of the book, about 34 pages, is devoted to the hot wire microphone and its application to sound location of aircraft and guns, and a further 10 pages to the acoustical design of aircraft sound locators. A discussion of the siren takes thirteen pages. Although aircraft noise and its measurement is considered in eight and a half pages there is little if anything on other noises and their measurement and suppression. Certain subjects are introduced in an unusual way, e.g. the Doppler effect through aircraft sounds without further mention of application, and the Whispering Gallery effect through observations on large sound reflectors for the detection of approaching aircraft. Some of the treatment appears to be unnecessarily dated, e.g. the recorded frequency spectrum of gramophone recordings is quoted as 50–5,000 c/s. and the last reference given is 1926. No mention is made of the recent improvements in wire and tape magnetic recording. Readers may argue however that such details are unnecessary in a textbook of this standard since they do not add to the fundamental concepts.

An erroneous impression may be given by certain statements such as: on p. 93, the uses of the tuning fork as a present standard of frequency for radio and clock control; on p. 242, the highly absorbent walls of broadcast studios; and on p. 134, the description of the mechanics of the electronic musical instruments. There seems to be some misunderstanding on the relative positions of the sound and vision gates on film projectors.

On musical questions, scale, consonance, etc., tend to be considered as a result of numerical arrangement. Numerical approach has interest but is mostly sterile.

The last chapter in the book, viz. "The Acoustics of the Concert Hall" could be better titled. Only a quarter of the printed matter refers to this question.

In a future edition an improved text could be obtained by a rearrangement of the material, with the deletion or considerable reduction of certain sections and the incorporation of further diagrams and photographs on the pure acoustical side. The addition of some thought-provoking problems at the end of chapters is also suggested.

H. D. P.

PUBLICATION RECEIVED

High Vacuum Technology, by A. S. D. BARRETT.

Abridged version of the 23rd Anniversary Address given to the King's College Engineering Society at the Institution of Mechanical Engineers on 10th March 1949. (London: W. Edwards & Co. Ltd.)

CONTENTS FOR SECTION A

PAGE

Dr. M. W. FEAST. The Schumann-Runge O_2 Emission Bands in the Region 3100 Å.-2500 Å.	549
Dr. M. W. FEAST. New O_2^+ Second Negative Bands: A Note on O_3 and OII Emission Spectra	557
Dr. M. W. FEAST. The Spectra Emitted by the High Voltage Arc in Nitrogen, Hydrogen, Nitrogen-Hydrogen Mixtures and Ammonia	563
Dr. M. W. FEAST. Rotational Analysis of the (1, 0) Band of the N_2 First Positive System	568
Dr. H. D. EVANS. An Absorption Comparison of the β -Particle Spectra of $^{207}_{81}AcC''$ (allowed), $^{218}_{88}RaE$ (second forbidden) and 3.5 yr.- $^{204}_{81}Tl$ (third forbidden)	575
Mr. G. R. BALDOCK. Excited Electronic Levels in Conjugated Molecules—V: A Valence Bond Estimation of Energy Levels in Aromatic Hydrocarbon Molecules	585
Mr. J. W. COOK, Miss R. SCHOENTAL and Mr. E. J. Y. SCOTT. Relation between Bond Structure and the Longest Ultra-violet Absorption Band of Polycyclic Aromatic Hydrocarbons	592
Mr. S. T. BUTLER. The Scattering of High Energy Charged Particles by Thin Foils of Matter	599
Dr. M. G. NOOH and Mr. S. R. HADDARA. Penetrating Showers at High Altitude.	606
Dr. J. H. VAN DER MERWE. On the Stresses and Energies associated with Inter-Crystalline Boundaries	616
Mr. R. B. DINGLE. The Theory of the Propagation of First and Second Sound in Helium II.—Energy Theorems and Irreversible Processes	638
Mr. J. D. LAWSON. The Angular Distribution of Synchrotron Target Radiation: A Preliminary Experimental Study	653
Letters to the Editor:	
Mr. D. J. LEES and Mr. L. H. METCALFE. Measurement of Polar Diagram of Synchrotron Gamma Radiation	661
Mr. F. K. GOWARD and Mr. J. J. WILKINS. Identification of Photo-Disintegration Stars in Nuclear Emulsions	662
Mr. J. J. WILKINS and Mr. F. K. GOWARD. Identification of Nitrogen Photo-Disintegration Stars in Nuclear Emulsions	663
Dr. A. P. FRENCH and Dr. P. B. TREACY. Alpha-Particles from $^{27}Al + D$	665
Dr. A. P. FRENCH, Dr. P. MEYER and Dr. P. B. TREACY. Alpha-Particles from ^{19}F Bombarded with Deuterons	666
Dr. JOAN M. FREEMAN. The Energy-Release in some ($p\alpha$) Reactions in Light Nuclei	668
Prof. E. A. STEWARDSON and Mr. H. F. ZANDY. The L-M and K-M Discrepancies in the Rare Earths	670
Mr. R. W. K. HONEYCOMBE. Inhomogeneity of Deformation in Metal Single Crystals	672
Dr. M. AFAF. Singlet System B of ZrO	674
Reviews of Books	675
Contents for Section B	678
Abstracts for Section B	679

ABSTRACTS FOR SECTION A

The Schumann-Runge O₂ Emission Bands in the Region 3100 Å.-2500 Å., by M. W. FEAST.

ABSTRACT. Rotational and vibrational analyses of the emission Schumann-Runge O₂ bands in the region 3100 Å. to 2500 Å. are presented. Tables of wave-numbers are given for the (1, 12), (0, 11), (1, 11), (0, 10), (1, 10), (1, 9), (2, 9), (1, 8), (2, 8), (2, 7) bands. The origins of the bands, the rotational differences, and constants for the levels $v''=7$ to 11 are given. Three bands, found in a high voltage arc in oxygen between platinum electrodes, are attributed to PtO.

New O₂⁺ Second Negative Bands: A Note on O₃ and OII Emission Spectra, by M. W. FEAST.

ABSTRACT. The spectrum of an electrodeless discharge in pure oxygen has been studied. New bands of the O₂⁺ Second Negative system have been found and the intensity distribution in this system is discussed; subsidiary parabolae as well as the main Franck-Condon parabola are obtained as is predicted by the theory. Emission bands reported by Johnson as due to O₃ are due to O₂⁺ and OII. Some abnormal intensities of OII lines excited in the electrodeless discharge are noted. The electrostatic and electromagnetic types of excitation are briefly compared.

The Spectra Emitted by the High Voltage Arc in Nitrogen, Hydrogen, Nitrogen-Hydrogen Mixtures and Ammonia, by M. W. FEAST.

ABSTRACT. The spectra emitted by the high voltage arc in nitrogen, hydrogen, hydrogen-nitrogen mixtures and ammonia have been studied.

The appearance of the N₂ Second Positive bands at various nitrogen gas pressures is discussed and the divergence of rotational and vibrational temperatures is noted. Traces of oxygen greatly reduce the strength of the N₂⁺ bands emitted by the arc in nitrogen. In N₂-H₂ mixtures the NH 3360 Å. band is found to be emitted strongly only when hydrogen is present at a very low partial pressure. An attempt is made to explain the various observations.

Rotational Analysis of the (1, 0) Band N₂ First Positive System, by M. W. FEAST.

ABSTRACT. The (1, 0) band of the N₂ First Positive system has been photographed at a dispersion of about 1.2 cm⁻¹/mm. The wave numbers and quantum numbers of the lines, and the rotational term differences in the upper and lower states are given. The spin tripling in the ³Σ state and the Δ doubling in the ³Π state are evaluated.

An Absorption Comparison of the β-Particle Spectra of ²⁰⁷AcC'' (allowed), ²¹⁰RaE (second forbidden) and 3.5 yr.-²⁰⁴Tl (third forbidden), by H. D. EVANS.

ABSTRACT. Experiments have been described in which a direct comparison is made by an absorption method of the β-particle spectra of actinium C'', radium E and 3.5 yr. thallium representing allowed, second forbidden and third forbidden transitions respectively.

New values of 623 ± 4 mg/cm² and 300 ± 3 mg/cm² are proposed for the absorption limits in aluminium of the β particles emitted by actinium C'' and thallium and it is shown that, of the three elements, radium E emits most and actinium C'' least low energy electrons.

An attempt has been made to derive energy distributions from the absorption curves, but the results for radium E depart markedly from those obtained with magnetic spectrographs. The general findings, however, agree with those of other workers who have attempted the same procedure.

The formula used in the derivation of the energy distribution includes an approximation which necessarily leads to an apparent deficiency of low energy particles, but the accuracy of the data is not sufficient to permit the use of a better approximation. Strictly for purposes of comparison, however, the method has obvious applications.

Excited Electronic Levels in Conjugated Molecules—V: A Valence Bond Estimation of Energy Levels in Aromatic Hydrocarbon Molecules, by G. R. BALDOCK.

ABSTRACT. This note is concerned with the calculation of the energy of the ground state and the lowest excited state for hydrocarbons of the condensed ring type. The difference between these two energies is the energy of the first electronic transition. The normal valence bond method is used, taking only the Kekulé structures into consideration. The results are compared with the experimental absorption wavelengths, and are in good agreement, giving a value of the exchange integral α between -1.8 and -2 electron volts.

Relation between Bond Structure and the Longest Ultra-violet Absorption Band of Polycyclic Aromatic Hydrocarbons, by J. W. COOK, R. SCHOENTAL and E. J. Y. SCOTT.

ABSTRACT. A comparison has been made between the values expressed in wave numbers of the longest ultra-violet absorption bands of a series of polycyclic aromatic hydrocarbons and the minimum number of quinonoid rings contained in their molecules. It is found that the hydrocarbons fall into series of groups related to this number of quinonoid rings.

The Scattering of High Energy Charged Particles by Thin Foils of Matter, by S. T. BUTLER.

ABSTRACT. The distribution of high energy charged particles scattered by a thin foil has been investigated and compared with the formula for single scattering. It is shown that at sufficiently large angles the distribution approaches that for single scattering and can be expressed in the form of a rapidly convergent series of which the first term is identical with the single scattering formula. The remaining terms of the series form an accurate expression for the deviation from true single scattering. The simplified case of projected scattering has also been considered.

Penetrating Showers at High Altitude, by M. G. NOOH AND S. R. HADDARA.

ABSTRACT. A study has been made of cloud chamber photographs of penetrating cosmic-ray showers, originating in lead placed over the cloud chamber and in a lead plate within the cloud chamber, obtained at altitude 3,572 m. in a magnetic field of about 1,000 gauss.

The use of well-separated successive lead layers allowed the association of penetrating events to be demonstrated, and there is some evidence of a general reduction of energy involved in events in the lower lead layers as compared with those in the top layer. Mesons, protons and particles of higher charge were identified in the penetrating showers, and a mean range for non-coulomb scattering of shower particles is given. Evidence is also given of the apparent starting and stopping of charged particles in the lead plate at the centre of the cloud chamber.

On the Stresses and Energies associated with Inter-Crystalline Boundaries, by J. H. VAN DER MERWE.

ABSTRACT. Models, largely based on the assumptions introduced by Peierls and Nabarro in dealing with a single dislocation, are used in calculations on three types of intercrystalline boundaries, namely, (I) a boundary due to a difference of atomic spacing, (II) a twist boundary, and (III) a symmetrical tilt boundary. With these models the resolution of the boundary into a sequence of dislocations is a natural consequence of the analysis, which also yields the expressions for the stresses, atomic displacements and energies associated with the boundary, as functions of the angle of tilt, etc. By allowing the distance between dislocations to tend to infinity, these expressions reduce to the corresponding ones for single dislocations. The outstanding feature of the interfacial energy is that it increases initially very rapidly with the angle of tilt, etc. An application of the results to the theory of oriented overgrowths, developed by Frank and the present author, is described. The validity of the assumptions and approximations involved and the advantages of the treatment as compared with those of other writers are discussed.

The Theory of the Propagation of First and Second Sound in Helium II.—Energy Theorems and Irreversible Processes, by R. B. DINGLE.

ABSTRACT. The first part of this paper contains a discussion of the conditions determining whether sound waves in helium II are propagated isothermally or adiabatically. In the remainder of this part an expression is found for the energy flux from a closed region, and the impedance concept is applied to waves of second sound.

The second part of the paper is devoted to the study of the effects of irreversible processes on the propagation of waves of first and second sound. The irreversible processes discussed are those due to viscosity and thermal conduction, both in an extended fluid and in a narrow tube.

The Angular Distribution of Synchrotron Target Radiation: A Preliminary Experimental Study, by J. D. LAWSON.

ABSTRACT. The angular distribution of radiation from a synchrotron has been measured with an ionization chamber for a number of different target materials and thicknesses, and the results have been compared with a simple theory in which the radiation from a parallel beam of electrons striking a thin plate has been calculated. Detailed comparison with theory is not possible, because the precise way in which the electrons hit the target cannot be determined, nevertheless qualitative explanations have been found for most of the effects observed. The most important of these is the relative insensitivity of forward intensity and beam width to target size and material.

THE PROCEEDINGS OF THE PHYSICAL SOCIETY

Section B

VOL. 63, PART 7

1 July 1950

No. 367 B

The Sliding Surface*

By G. I. FINCH

34th Guthrie Lecture, delivered 3rd March 1950

§ 1. INTRODUCTION

LIFE is possible because some assemblages of matter can store sunlight as chemical energy. Other organisms, unable to do this directly, absorb these stores to supply their own energy needs; and some, like man, prey on both to live. Man alone supplements the body's production of physical energy by using inanimate devices to unlock the stores of energy latent in the remnants of past life. The key is the machine, which works because smooth surfaces can slide over each other. The sliding surface is, therefore, fundamental to our modern way of life.

The sliding surface was known long before the machine; away back in antiquity man had used it to ease the burden on his back. He had also learnt the value of a generous supply of some slippery substance. So when the machine did arrive the engineer saw to it that its sliding surfaces were lubricated; so well, indeed, that they were kept apart by an oil layer many hundreds or even thousands of molecules thick. Nowadays, however, sliding surfaces are usually so heavily pressed together that the lubricant film cannot be more than a few molecules thick, and is often penetrated or disrupted so that the sliding surfaces touch in places.

Thick-layer, or, as it is generally known, hydrodynamic lubrication (Reynolds 1886) favours easy sliding without wear other than that due to corrosion and erosion, but with boundary (or molecular film) conditions† of lubrication, where the sliding surfaces often come into contact, suppression of wear becomes the main problem. Thus, in hydrodynamic lubrication chief interest is directed to the lubricant, but in boundary lubrication it is the sliding surface which claims first attention. What causes rubbing surfaces to wear and how do they wear? What peculiarity in the surface helps to reduce wear and how does it work? How do the surfaces slide when they touch each other? What is it about the surfaces that helps them to slide and how does it help? We must first seek the answers to questions like these before turning to the study of the lubricant boundary film.

As far back as the eighteenth century Coulomb (1781) thought that resistance to sliding was due to the interlocking of surface asperities, and that the fracture of these during sliding was responsible for wear. Ewing (1912) and Hardy and

* This Lecture will also be published in Section A in August.

† The term 'boundary conditions' was first used by Reynolds.

Hardy (1919), on the other hand, believed that the interplay of surface cohesive forces was the dominant factor. While Coulomb's view may once have served a useful purpose, it is completely at variance with modern engineering experience. Furthermore, the results of critical surface studies by Beilby (1921), Hardy (1936), Finch (1937), Bowden and Tabor (1939) and Parker and Hatch (1950) do not support Coulomb's hypothesis but are readily interpreted in terms of surface cohesive forces.

The forces which produce cohesion within a block of metal must also act at the surface of contact between two blocks; they should, therefore, resist stress as if the blocks were one. There are, indeed, facts which support this conclusion. Slip gauges left 'wrung' together eventually seize and are damaged when separated. Hardy and Hardy (1919) found that some solids seized together on contact. Fifteen years ago I showed—and it must have been known long before then—that mica cleavage surfaces could adhere firmly when brought together (Demonstration). But experience also shows that surfaces can, and often do, slide easily over each other, even in the absence of lubricating oil.

Although a solid exerts a cohesive force on an atom beyond its surface, the intensity of the force decreases rapidly with distance, to become negligible a few atomic radii away. Contact between two surfaces, therefore, implies approach to within this distance. It is possible, by cleaving certain mineral crystals, to produce atomically smooth plane surfaces of considerable area, but otherwise such a degree of smoothness cannot in general be achieved or even approached. Hence, when two nominally plane surfaces formed, let us suppose, on two blocks of the same metal are placed together, contact can occur only at three points. If the surfaces are now loaded, stress at the points of contact must exceed the elastic limit. Plastic flow will then enlarge the points of contact to areas, and fresh points of contact expanding to still more areas will be established, until the load is borne without further deformation of the material.* A considerable effort should now be required to pull or shear the blocks apart because, at the areas of contact, they should have become united. Experiment, however, gives a different result. Thus, mica cleavage surfaces which have been exposed to the atmosphere for some time, or momentarily immersed in water, will no longer adhere, no matter how hard they are forced together (Demonstration). If, instead of being 'wrung', two slip gauges are merely pressed together, they do not stick but fall apart when the pressure is released (Demonstration). Mechanically or anodically polished surfaces of copper, iron, silver, tin and other metals behave similarly, provided the pressure does not cause general plastic flow at the surfaces (Demonstration). An atomic degree of smoothness, superimposed, it is true, on a large-scale waviness of small amplitude, may be imparted to a face of a copper single-crystal by anodic polishing; nevertheless, a degree of pressure short of that causing permanent distortion will not even partially weld two such surfaces together (Demonstration). In this case the undimmed clarity of the electron diffraction pattern shows that any adsorbed layer cannot be more than two or three atoms thick (Figure 1).†

* I frame these conclusions on the basis of Amontons' (1699) experimental law according to which the coefficient of friction is independent of the area of the sliding surfaces. The validity of this law has often been confirmed. The reasons for apparent exceptions are well understood; they do not affect the issue.

† Figures 1 and 3–20 are photographs and are printed together opposite p. 472.

It may therefore be inferred that true contact is not as a rule established when even atomically smooth solid surfaces are pressed together, but that they are kept apart by adsorbed surface layers which, while satisfying the cohesive forces at the surfaces, exert little or no attraction towards each other. The remarkable resistance of these layers to penetration appears to be independent of their nature (if it is assumed that this varies with the metal) and determined mainly by the physical properties of the metal, general plastic flow of which might be expected to rupture such a layer by extending its surface, and thus allowing welding to occur through direct metal-to-metal contact. It will be evident that this adsorbed layer is of vital importance. Without it, a pair of surfaces at rest under load would seize together over an area proportional to the imposed load. It cannot be supposed that welds are formed only to be broken when, on release of pressure, the surfaces are forced apart by a recovery of the metal about those areas where the elastic limit has not been exceeded; for the metal in such welds, being itself compressed to the elastic limit, would have the greater range of recovery on release of the pressure.

When the surfaces are not only pressed together but forced to slide past each other, the effectiveness of the adsorbed layers in preventing metal-to-metal contact varies greatly according to the nature of the metal. Anodically polished single-crystal copper or aluminium surfaces at first slide then suddenly lock firmly together, even with light loading (Demonstration). Similarly prepared polycrystalline surfaces of these metals also seize quickly, though not quite so readily as single crystals (Demonstration); the corresponding mechanically polished surfaces slide more easily and are much more resistant to seizure (Demonstration). Anodically polished polycrystalline silver or iron single-crystal pairs slide freely and must be heavily loaded before signs of seizure appear (Demonstration), and similar experiments with combinations of anodically polished single-crystal and polycrystalline surfaces of various metals reveal further wide differences in the ease or otherwise with which sliding and particularly seizure take place. But whether this is to be attributed to the physical properties of the metal, as in the case of the loaded surface pair at rest, or whether the nature of the adsorbed layer is also significant, remains to be seen.

The existence of these layers has long been known. Faraday (1838) and De la Rive (1839) recognized their importance in heterogeneous catalysis, and they have since been much studied by the chemist. The physicist's interest seems to have been first fully aroused by their effect on thermionic and photoelectric emission (Langmuir 1916). The engineer, so long accustomed to hydrodynamic lubrication, has been slower to appreciate their value.

The layer is difficult to remove from a metal surface. It can be done by scraping or evaporating *in vacuo* and so exposing a fresh surface, or by heating to a high temperature *in vacuo* or in the presence of atomic hydrogen; but the layer re-forms spontaneously and rapidly, even at room temperature, on exposure to air. These facts suggest that the adsorption is chemical rather than physical, and that the layer is an oxide of the metal. This has been amply confirmed by electron diffraction.

Although oxidation is the immediate result of the exposure of a fresh metal surface to the atmosphere, the microbalance reveals that a moisture film of considerable thickness (as much as 80 molecular layers and possibly more) may be adsorbed

on the oxide surface. When fresh, this layer is as a rule not strongly held and is removed by the electron beam in the evacuated diffraction camera.

§ 2. THE SLIDING ELEMENT

It is clear that the sliding element of the modern machine must be a highly complex system. Actually its simplest form comprises at least nine distinct components (Figure 2). These are (i) and (ii) the metallic backings supporting (iii) and (iv) the pair of metal surface layers or so-called bearing surfaces which, in turn, carry (v) and (vi) their appropriate oxide layers each with its (vii) and (viii) adsorbed boundary layer of lubricant, and (ix) an oil film of varying thickness held between the boundary layers and more or less charged with debris, moisture and gas bubbles. The components in each pair may differ widely in nature and properties, even in the case of the two lubricant boundary layers. Furthermore,

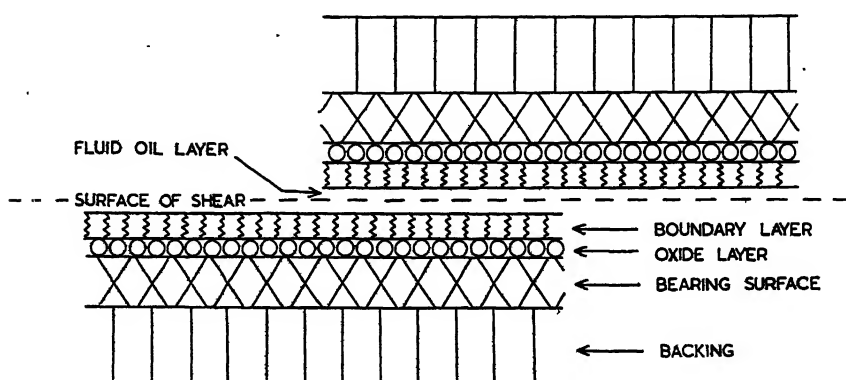


Figure 2. Diagram of the sliding element.

a change in any one component may also affect the behaviour of the others. It is this astonishing complexity in what at first sight appears to be a simple system that explains why, in the past, research has been for the most part directed towards the special circumstance of practice, rather than to a fundamental study from which broad generalizations might be drawn. Clearly, the physicist's approach to the problem must be to break down the sliding element into its simplest components and to examine each in turn, first by itself and then in relation to the others.

§ 3. THE BACKING

The backing is rigidly secured, directly or indirectly, to an appropriate member of the machine. Its function is to support the bearing surface so that its motion is constrained to the desired degree or degrees of freedom with respect to the other sliding member. In the case of a journal bearing, the force tending to rotate the backing in its housing may be considerable and is given by the product of the coefficient of friction between the sliding members and the load. Since the polar distribution of the load may vary periodically between wide extremes, particularly in the case of the internal combustion engine, the backing tends to oscillate in the housing. This may lead to serious fretting corrosion and loss of fit.

Backing and bearing surface may be composed of the same metal, as in the shaft of a journal bearing; or the metals may differ, as in the liner of the housing within which the shaft rotates. As the properties of the backing may affect the behaviour of the bearing surface, it will be convenient to defer its further discussion.

§ 4. THE BEARING SURFACE

The bearing surface is developed in practice as a plane or surface of revolution by abrasion processes such as cutting, grinding, honing and lapping. These operations reduce the crystal size of the metal at the surface, and often to a considerable depth below it (Figure 3). This 'work-hardening' can seriously impair the efficiency of the bearing surface, and for this reason the more severe operations of cutting and grinding should be followed by annealing to restore the uniformity of crystal size and relieve stress (Finch 1945). The texture of the finished surface is, in terms of atomic dimensions, exceedingly rough and made up of rows of steep hills and valleys, each with smaller irregularities and lying roughly parallel to the direction of cutting. The ridge crests are seen by electron diffraction to be crystalline (Figure 4).

General experience has shown that the fresh bearing surface must not be subjected to full loads and sliding speeds until it has been 'run-in' under light loads with copious supplies of lubricant. If the protuberances are not gradually reduced, as they are by 'running-in', the crystal structure will be broken up to such a depth below the surface as permanently to impair the sliding and resistance to wear of the surface. The effect of 'running-in' may be explained briefly as follows. Consider, first, bearing surfaces of similar metal. When, through breakdown of the oil and oxide layers, metal-to-metal contact is established between opposing high spots of the bearing surfaces, a weld is formed. As sliding motion proceeds the weld is stressed and eventually ruptured. If the crystal structure in the surfaces has not previously been broken up and therefore hardened, the weld will shear at or near one of the surfaces. On the other hand, if, owing to previous unduly severe conditions of machining or sliding, work-hardening has occurred to some depth below the surface, the weld, instead of rupturing virtually in the plane of sliding, will continue to adhere to one surface and pull out work-hardened metal to a considerable depth from the other (Figure 5). The protruding metal will then score and break up the crystal structure in this surface until it has been worn away; or it may break off and roll between the surfaces and damage them by crushing (Figure 6).

It is now known that sliding causes a profound and important change, not only in the texture but also in the structure of the metal surface. At first, rubbing of the sharper asperities on the opposing fresh bearing surfaces results largely in wear and the production of debris. As the surface becomes smoother, however, wear decreases and much of the metal removed from the hills 'flows', as Beilby (1921) puts it, into the valleys. Thus 'running-in' levels off the mountain ridges and fills the valleys to leave a texture like undulating hills. This smoothing is accompanied by a remarkable change in structure (Finch, Quarrell and Wilman 1935). Not only are the crystals broken up to a depth below the surface depending on the severity of 'running-in', but the surface layer structure becomes amorphous (Figure 7), as if 'flowing' consisted of the transfer of molten metal from one part of the surface to another where, having escaped the polishing action, it freezes too rapidly to crystallize. The thickness of this so-called Beilby layer depends, like that of the work-hardened crystalline zone beneath, on the severity of the 'running-in' conditions. Layers of only about fifteen atoms thick have been produced (Hopkins 1935) and others about 100 atoms thick have been successfully detached from their substrates (Cochrane 1938); but a vigorous polishing results in a much thicker layer (Finch, Quarrell and Roebuck 1934).

The following experiments will show how great is the protection afforded to the oxide layer by the support of even a very thin Beilby layer. Severe welding and extensive damage occurs when two anodically polished copper single-crystal surfaces are slid past each other (Figure 8) (Demonstration). The behaviour of a similar pair of surfaces, on one of which, however, a very thin Beilby layer has been formed by light polishing with rouge, is very different. Much heavier pressure is needed to cause welding, and on forcible separation of the surfaces metal is pulled out from the polished surface only to a shallow depth (Figure 9) (Demonstration). Thus the Beilby layer performs important functions in sliding. Its comparative smoothness favours a wide distribution of load where 'high spots' meet as the bearing surfaces slide past each other; it is more rigid than the underlying crystalline metal and thus better able to support the oxide layer and preserve it from injury; and a cold weld, with its tendency to pull out near the zone where the amorphous merges into the undisturbed crystalline structure, will do least damage if the Beilby layer is thin.

The relative amounts of metal worn away (and thus lost to the surface) and metal which is merely displaced to form the Beilby layer, are affected by the direction of the ridge and valley lines produced in machining in relation to that of sliding (Finch and Zahoorbux 1937). It has been found that the heaviest wear occurs when these lines in both bearing surfaces lie in the direction of sliding, and that wear is least when they are at right angles to the sliding direction. If the lines on one surface are parallel to the direction of sliding, while those on the other are at right angles to it, then wear of the former surface is high. The explanation is simple. When the lines in both surfaces are parallel to the sliding direction, the metal, severely stressed in localized contacts has farther to flow before it can escape and congeal beyond the pressure zone, and is thus more likely to form unattached detritus than when the ridges and valleys lie at right angles to the sliding. The importance of the 'run' of tooling tracks in relation to the sliding direction is now appreciated in engineering practice.

We have seen that when two nominally plane surfaces are pressed together the total cross-sectional area of the resulting intermetallic contact is proportional to the load. The force required to start sliding is therefore also proportional to the load; its value per unit load is known as the coefficient of static friction. To maintain sliding, work must be done. The necessary force is again proportional to the load, provided shear occurs at the interface and the strength of the metal in the contacts is unchanged. Under such conditions the dynamic and static coefficients of friction should be equal. Micrographic examination shows, however, that the cold welds formed on contact between, for example, single-crystal copper surfaces are strain-hardened as sliding proceeds, and also that the surrounding metal is similarly affected to a considerable depth (Figures 5 and 8). Hence, instead of the welds shearing at the interface, metal is pulled out bodily from below one surface to be dragged along and ruck and score the other (Figure 5). Under such conditions the dynamic coefficient exceeds the static. In normal practical conditions of sliding, of course, the reverse is the case, as the oxide and lubricant films together greatly limit intermetallic contact during sliding and may even in some cases inhibit it. Nevertheless, as a rule, spasmodic intermetallic contact does occur, and its contribution to the total resistance to sliding may be called intermetallic friction. Its most undesirable feature is that it causes wear.

A sliding element in which the intermetallic component of the total friction is considerable is therefore unsatisfactory.

So far we have considered bearing surface pairs of the same metal. In practice, however, where sliding speeds are high, this combination is avoided except in the case of cast iron. Otherwise the engineer's empirical rule is that the mating bearing surfaces should consist of metals of widely differing melting points. It can, indeed, be inferred from Bowden and Ridler's (1936) measurements of the temperatures attained at metal contacts in sliding that localized fusion followed by hot welding should occur more readily with like than with unlike metals. The matter, however, goes deeper than this. Solid mercury and gold would make an unsatisfactory pair of bearing surfaces.

Now, we have seen that in addition to hot welding due to melting of rubbing high spots on both bearing surfaces, cold welding also occurs on contact of the solid metals. Indeed, under the conditions of the experiments I showed you, the friction and surface damage was entirely due to cold welding and its after effects. Even with like metals, it is only under severe sliding conditions and more particularly at high speeds that hot welding becomes important. It will be evident, therefore, that in selecting metals for a bearing surface pair, one criterion is the intensity of the cohesive force across the interface of intermetallic contact, rather than disparity in their melting points. If this interfacial cohesion is less than that within the weaker of the two metals, intermetallic contacts will shear in the interface. The other criterion, which concerns the nature of the oxide layer, will be discussed later.

The problem of estimating this cohesive force from a knowledge of the structures of the outer electron shells of the atoms of which the contacting surfaces are composed has not been solved. However, a study of the coordination relationships between epitaxial metallic deposits and metallic single-crystal substrates is now being carried out in my laboratory (Finch and Sun 1936, Finch and Williams 1937, Finch, Wilman and Yang 1947); this seems to offer a promising line of experimental approach to the problem. In the meantime, as a rough and ready rule, we may take it that the more rigid the metals of a pair and the higher their melting points, the less will be the tendency to form *cold welds*.*

It would, I think, be fair to say that, as in the development of most aspects of the sliding surface, the choice of metals for bearing surface pairs has been empirical. The modern journal bearing practice is to use pairs in which one metal is hard and has a high melting point and the other is relatively soft and fusible. In the sliding of such combinations new phenomena appear.

Bowden and Hughes (1937) showed that when certain solids were rubbed together it was not the harder surface which caused the softer to flow and thus be polished, but that the more fusible surface was polished by the more refractory one, irrespective of their relative hardnesses at room temperature. In practice, however, sliding surfaces may well behave differently. For example, a type of journal bearing in common use consists of a superficially hardened alloy steel

* It would be virtually impossible to put together even two perfectly plane complementary single-crystal surfaces on a pair of blocks of the same metal in such a way as to restore the original atom coordination across the interface. The higher the mobility of the atoms, however, the greater is the chance of the atoms in the opposing faces moving into and being retained in sites of lowest potential energy. This is akin to a process of recrystallization whereby the two blocks eventually become welded together into a single crystal.

shaft rotating against a bearing surface of a rather soft tin-base alloy melting at about 250°C . Since the steel melts at roughly $1,400^{\circ}\text{C}$. the alloy surface becomes polished as one expects; what is surprising, however, is that the hardened surface of the shaft also flows and is polished as running-in of the bearing proceeds. Similar observations have been made in reciprocating sliding. Indeed, in all practical cases so far examined this mutual polishing of sliding metal surfaces has proved to be the rule, irrespective of the difference in hardness or fusibility of the bearing surface metals. The explanation is to be sought in the nature and properties, including the geometry, of the oxide layers and metals of the bearing surfaces, and will be discussed later.

Quite apart from the reduction in welding, and hence in friction and wear, further advantages can accrue from the use of a relatively fusible soft alloy sliding on a hard refractory bearing surface. Owing to the comparative ease with which the softer metal is plastically deformed, it was long thought necessary to restrict the load intensity. The further disadvantage of the low ductility and proneness to fracture of the customary rather thick strip of soft metal was overcome, at least in part, by welding to a massive backing of a metal of more suitable mechanical properties, which was then secured in an appropriate housing. The resulting bearing, however, was large and massive in relation to the applied load. Hopkins and Palm (1932) realized that, by taking advantage of the cohesion between the bearing surface and the backing metals, the softer metal could be made to withstand a load intensity far greater than hitherto thought practicable. The underlying principle is simple. The ability of a soft body to resist deformation when compressed between two hard plane surfaces increases as its thickness is reduced by plastic flow until, long before the soft material can be totally extruded, the hard metals themselves begin to flow and deform. For example, it is impossible to force out completely even an oil film from between two plane surfaces, no matter how heavy the load or how rigid the metal may be. This is due to the reinforcing of the cohesion between the oil molecules by their adhesion to the surfaces. In fact, the thinner the film, the more do its mechanical properties approximate to those of the solid substrate. Similarly, whilst a bearing metal layer of the once customary thickness can support only a relatively low load intensity, a much thinner layer can carry a far heavier load without being permanently deformed. The thickness of this layer need be no greater than that corresponding to the maximum wear loss which can be tolerated before the fit becomes so slack as to render the bearing useless. For example, the 'big-end' crankshaft bearing of an internal combustion engine, which previously consisted of a 5/100 in. thick tin-base alloy fused on to a massive bronze backing, has now been replaced by a 5/1,000 in. thick layer on a thin steel strip, the weights of the liners being 218 and 60 gm. respectively (Exhibits).

The tin-base type of bearing metal alloy, though relatively soft, is too hard to be scratched with the finger nail. This form of alloy, though extremely wear-resisting, causes considerable wear of the hardened steel surface against which it slides. As we shall see later, this is due partly to the softness of the alloy and partly to the nature of its oxide layer. There is little evidence to support the view that embedded detritus particles are the cause. A further notable advance therefore was made by the recent introduction of bearing metals whose surfaces cause very little wear of the steel bearing surface. Lead and indium, the chief metals of this kind, are so soft that they and their alloys can be scratched with the finger nail.

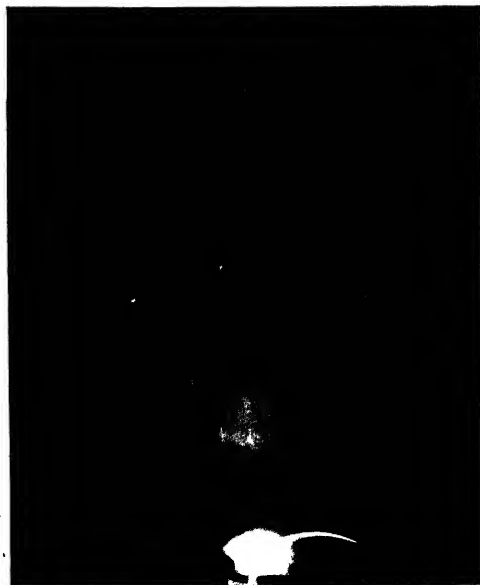


Figure 1. Electron-diffraction pattern from a copper single-crystal electropolished cube face; beam along cube edge. The sharp Kikuchi lines and the narrow circular Laue zone show that the surface is atomically smooth and of quite remarkable lattice perfection. The clarity in detail near the shadow edge is such as to show that the oxide layer cannot be more than two or three molecules thick.



Figure 4. α -iron pattern from the freshly honed surface of an austenitic steel cylinder liner.

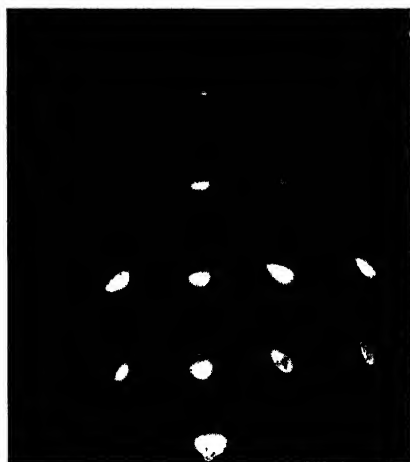


Figure 3. An iron single-crystal (110) face was lightly honed along a [110] direction and then etched to expose the sub-surface damage. The pattern (beam along [001]) reveals that the original single-crystal structure has been distorted by a process of rotational slip (Wilman 1950) on (001) about a [001] axis. This slipping represents the initial stage in crystal break-up. Note that the fragments are still coherent across the planes of rotational slip.

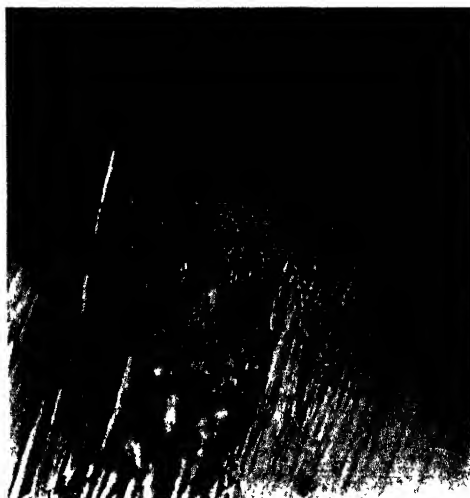


Figure 5. Scoring damage due to welding between a pair of copper single-crystal sliding surfaces. 450 \times .



Figure 6. Track of a broken-off weld particle rolling on a copper single-crystal surface. 'Formvar' replica; gold shadowed. 10,000 \times .



Figure 7. Mechanically polished copper surface. The pattern consists of two ill-defined broad haloes. Patterns of this kind are obtained from glass, fused quartz, cellulose, and from mechanically polished metal surfaces in general.

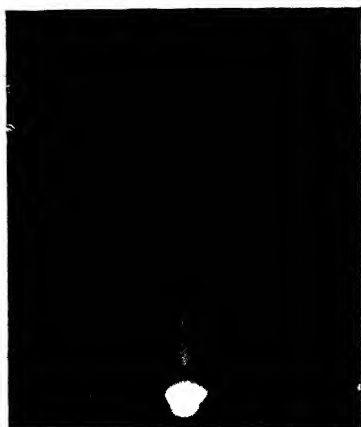


Figure 10. Pattern from an austenitic steel internal combustion engine liner after 120 hours' running. The halo pattern reveals the presence of amorphous iron and an amorphous iron oxide. The faint streak above the undeflected beam spot is due to a graphite film the crystals of which are orientated with their cleavage planes parallel to the liner surface.



Figure 8. Surface damage resulting from weld formation between a pair of copper single-crystal surfaces. Note widespread regions of deformation bands formed ahead of and at the sides of the main damage area. 450 \times .



Figure 9. The weld damage is greatly restricted when one of the copper single-crystal surfaces is first lightly mechanically polished. The vertical lines are due to the polishing medium. 450 \times .

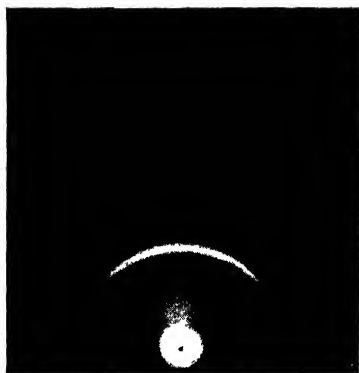


Figure 11. A scratched cast-iron piston-ring surface yields a pattern due to disorientated graphite crystals.

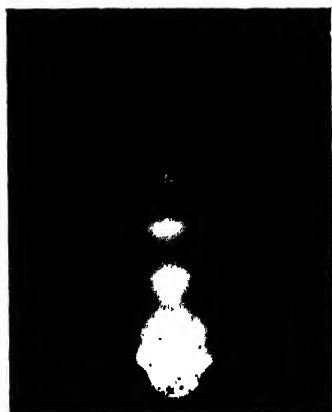


Figure 12. Light rubbing with the thumb suffices to orientate the graphite crystals with their cleavage planes parallel to the piston-ring surface. This pattern identifies the streak in Figure 10 as being due to graphite.

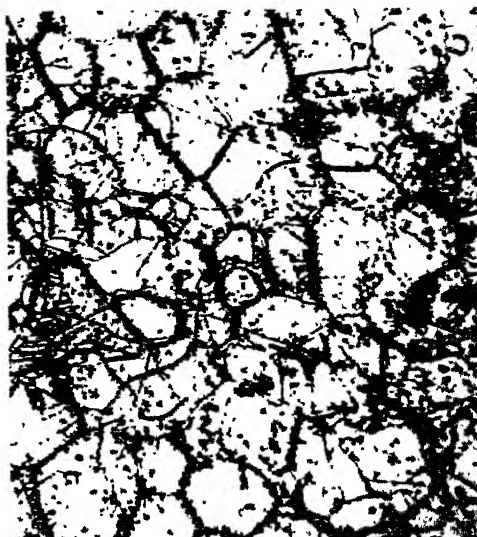


Figure 13. Polycrystalline copper surface after removing an oxide film. The crystal boundaries have been more attacked than the crystal surfaces.

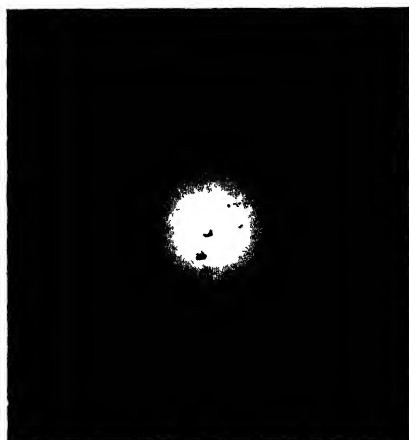


Figure 14. Amorphous iron oxide film removed from a mild steel surface.



Figure 15. $\gamma\text{-Fe}_2\text{O}_3$ formed by heating the amorphous oxide film of Figure 14.

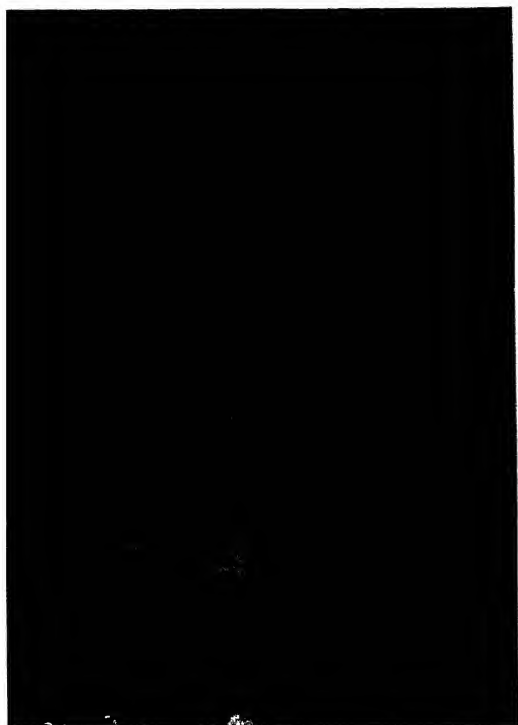


Figure 16. A copper single-crystal surface immediately after anodic polishing.



Figure 17. The same surface after 30 hours' exposure to air at room temperature becomes covered with an amorphous oxide layer which is thick enough ($>40 \text{ \AA}$) to obscure the single-crystal pattern.

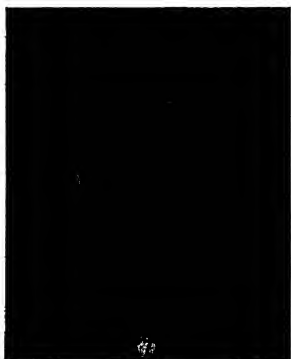


Figure 18. Cuprous oxide growing epitaxially on a heated macro-crystalline copper surface.



Figure 19. $\gamma\text{-Fe}_2\text{O}_3$ formed on a heated steel surface. The tendency is for needle-like crystals to form with $[110]$ axes normal to the surface.



Figure 20. General movement of carbon shot during descent of piston. (From Andrade and Fox (1949).)

Nevertheless, when bonded in a sufficiently thin layer to a steel backing, they can support unusually heavy loading. One such backing-bearing surface combination consists of a hard steel strip to which is bonded a thin copper-bronze alloy strip on which a layer of lead is electrodeposited. A thin 'flash' of indium is electrodeposited and later thermally diffused into the lead. The copper-bronze strip is necessary as, owing to the differences in crystal structure and lattice constants, it is difficult to deposit lead electrolytically on steel in such a way as to ensure a satisfactorily close co-ordination, and therefore good cohesion, between the atoms across the interface. The total thickness of the lead-indium layer is 1/1,000 in. The rate of wear of both bearing surfaces is so small in relation to that of the other sliding surfaces of the engine that this thickness is ample (Exhibits).

An important question relating to bearing surfaces which has so far received less attention than it undoubtedly merits is that of the effect of sliding on the surface composition of alloys. In the few cases in which this has been examined experimentally it has been found that polishing increases the proportion of one or more of the constituents in the surface of the alloy, and hence also affects the composition of the oxide layer. Thus polishing a chromium steel alloy produces a surface which is either greatly enriched in chromium content or is even pure chromium covered by a layer of chromium oxide (Vernon, Wormwell and Nurse 1944). A similar effect has been observed in this laboratory with magnesium-aluminium alloys (De Brouckère 1945), α -brass and bronze (Bryant 1936) which become richer in magnesium, copper and tin respectively. Such changes may well be significant, not only because of their possible effect on the surface hardness and fusibility, but also because they must determine the nature of the oxide layer.

The demands made on the reciprocating bearing surfaces in the cylinder of the internal combustion engine are in many respects much more severe than in the journal type of bearing. Surface temperatures are high, particularly near the combustion zone, and the pressure distribution on the piston crown is seldom uniform, especially at the start of the expansion stroke. Further, the change from dynamic to static conditions at the end of each stroke not only imposes unusual demands on the lubricant layers but also results in an uneven load distribution between the sliding surfaces. Finally, the reaction products cause corrosion wear (Williams 1937).

Nominally, the bearing surfaces are those of the cylinder wall and piston rings, but the piston itself at times also rubs against the cylinder wall. For reasons of mechanical strength, lightness and thermal conductivity, the piston may consist of an aluminium alloy with copper and some magnesium, though in diesel engines a cast-iron skirt with an alloy steel crown is more usual. The cylinder bearing surface is a cast-iron, austenitic steel or other ferrous alloy, sometimes with a coating of electrodeposited chromium. The piston rings, however, are always made of cast iron. Alloys with better mechanical properties have been tried as piston-ring metal in low-compression engines but their adequate lubrication presents difficulties. The graphite in the cast-iron ring helps greatly in the suppression of wear of both cylinder and ring (Finch, Quarrell and Wilman 1935, Finch and Whitmore 1938). Electron diffraction has shown not only that the 'run-in' surfaces of the cylinder and cast-iron piston ring are covered with an amorphous oxide layer (Figure 10) but has also revealed the presence of an exceedingly thin adsorbed graphite layer (Figures 11 and 12), the flakes of which are orientated with their cleavage planes parallel to the surface. It is probable that this layer acts as a

temporary lubricant and protection to the oxide layer when the oil boundary layer breaks down.*

The effect of direct sliding contact between the exceedingly hard corundum coating on the aluminium alloy piston and the oxide layer on the cylinder wall seems to be mitigated by the formation of magnesium spinel (Finch 1937). There is some evidence also that the rubbing of the corundum or spinel layers against the ferric oxide layer on the cylinder wall may lead to the formation of iron spinels (Thirsk and Whitmore 1940).

§ 5. THE OXIDE LAYER

In modern engineering practice, where frequent penetration of the lubricant boundary layer occurs, the part played by the oxide layers in facilitating sliding and preventing metal-to-metal contact is obviously of great importance. It is now known that most metals, if not indeed all, oxidize in dry air, however low the temperature. The primary oxide layer is formed very quickly, even at the temperature of liquid air, and then acts as a barrier against continued easy access of oxygen to the surface. Thus the subsequent rate of growth of the oxide layer decreases rapidly, and may become negligible at a thickness of the order of a few tens of atomic radii. Should further oxidation occur, it is very slow, even at room temperature, provided secondary effects, such as attack of the oxide films by adsorbed moisture, are avoided. Cabrera and Mott (1949) explain the mechanism of oxidation subsequent to the formation of the first oxygen layer in terms of the high contact potential gradient between the metal and the layer, which enables the metal ions to move through it without much help from thermal motion. There is, however, evidence which suggests that oxidation can also be due to oxygen gaining access to the metal through crystal boundaries. Thus, we have observed that an anodically polished single-crystal metal surface oxidizes more slowly than a similarly treated microcrystalline surface. Further, on removing with citric acid a relatively thick oxide film grown at 180°C. on anodically polished polycrystalline copper, it was seen that the crystal boundaries, barely visible before oxidation, had become strongly outlined and deepened (Figure 13).

The surface oxidation of metals is, in general, a highly energetic process. At low temperatures, therefore, the rapid arrival of oxygen molecules and the low mobility of its atoms over the surface may be expected to favour the growth of amorphous rather than crystalline layers. There are many examples of films having been grown on single-crystal substrates cooled to liquid air temperatures and, as far as I am aware, all have been found to be amorphous or so submicroscopically crystalline as to be indistinguishable from amorphous; yet the same

* It is commonly supposed that the crystals of layer-lattice structures such as mica, graphite and molybdenum sulphide, which are sometimes used as dry lubricants, act like orientated packs of playing cards. This cannot be so because, as I demonstrated earlier in this lecture, the layers can only be separated by cleaving in such a manner that the bonds between the layers are broken one by one. I also showed that, though freshly cleaved mica surfaces would adhere firmly on being first put together and could not be pulled apart by stress applied in a direction lying in the cleavage plane (i.e. attempting to rupture the bonds simultaneously), they lost this property on prolonged exposure to air or moisture. Thus the ability of layer-lattice cleavage surfaces to slide easily over each other is due to adsorbed layers. Layer-lattice crystals with their *c* axes parallel to, and crushed between, loaded sliding surfaces will be cleaved, but will also abrade the sliding surfaces. It follows that layer-lattice crystals in which the ratio of exposed cleavage plane area to thickness in the *c* direction is a maximum make the best form of solid lubricant. It is known from electron diffraction (Finch and Wilman 1936) and electron-microscopic studies (Rao, recent work in this laboratory) that the smaller the volume of a graphite crystal the higher is the value of the above ratio. A layer-lattice lubricant should, therefore, consist of crystals in the finest possible state of subdivision compatible with a maximum surface to thickness ratio. Sliding of the lubricated mechanism should not be relied on to break up the crystals, as this must result in wear.

substances deposited on the heated substrates have grown epitaxially. Even when a metal surface is oxidized at room temperature the oxide layer usually seems to be amorphous. This is certainly so in the cases of aluminium, chromium and iron (Figures 14 and 15); and recently Dr. Leu, working in my laboratory, has found that the oxide film formed on an anodically polished copper single-crystal face exposed to air at room-temperature is also amorphous (Figures 16 and 17).

On heating to above room temperature, amorphous oxide films differ widely in behaviour. Those on aluminium and chromium remain amorphous at temperatures up to about 600°C. and can grow in thickness only very slowly, if at all (Preston and Bircumshaw 1936, De Brouckère 1945). On copper, cuprous oxide grows epitaxially (Figure 18) and, at higher temperatures, becomes in turn superficially oxidized to cupric oxide (tenorite). These films, although coherent, are fairly easily permeable by both oxygen and copper ions, and thus do not give a high degree of protection against oxidation. Still less protection is given by the crystalline iron oxide films ($\text{FeO} \rightarrow \text{Fe}_3\text{O}_4 \rightarrow \text{Fe}_2\text{O}_3$) (Figure 19) formed under such conditions that the crystals have a marked tendency to grow outwards from the surface along an atom row of high atom population density. With other metals the crystalline oxide layer formed on heating is orientated in such a way that the crystals have in common a plane of high atom-population density parallel to the metal surface. They also cohere to each other and form a continuous film which hinders access of oxygen to the surface even when the metals are molten. Well-known examples are tin, zinc, antimony, bismuth, cadmium and lead. Oxide layers grown on heated metals when the rate of supply of oxygen is sufficiently reduced, are often epitaxial, i.e. the crystal orientation in the oxide layer is related to and determined by that of the metal surface. Such an oxide layer grown on a single crystal is itself virtually a single crystal. Epitaxial oxide growth can also occur during the oxidation of polycrystalline surfaces, even though the metal crystals have no common direction or directions of orientation. Then, however, the crystalline continuity of the deposited layer is further disturbed by the epitaxial reproduction of the irregularities which constitute the crystal boundaries.

Crystalline oxide layers on the sliding surface are probably of rare occurrence. Although a metal surface exposed momentarily through rubbing may be extremely hot, it is, as we know from the formation of the Beilby layer, disturbed and very rapidly cooled, so that there will be a tendency for any re-formed oxide layer to be amorphous. Many 'run-in' bearing surfaces have been examined by electron diffraction; their oxide layers have been found to be amorphous. The thickness of the amorphous films so far studied appears to be relatively uniform, so that the surface texture is a fair representation of that of the underlying metal. Indeed, the amorphous aluminium oxide layer has been used as a replica in the electron-microscopic study of the texture of aluminium surfaces.

The adhesion of the oxide film to its substrate can be expected to be good by reason of its nature, though pseudomorphic strain in the initial layers may not be without influence (Finch and Quarrell 1933, Cabrera and Mott 1949). As the film is amorphous, and therefore isotropic, it should be equally strong in all directions in the surface. Its cohesive strength in the case of iron and aluminium is shown by the way in which detached films, less than 100 Å. thick, of amorphous ferric or aluminium oxide can be lifted off a water surface on a wide-meshed (80/inch) gauze without tearing, a process which also demonstrates their flexibility.

We may now examine how the oxide layer is so effective in preventing metal-to-metal contact when two surfaces are pressed together. When the surfaces first come together they make a three-point contact. This does not mean that three pairs of points meet. The probability is rather that three asperities of small radii of curvature will lodge on surfaces of large radii. Such a contact pair may be represented by Figure 20 (Andrade and Fox 1949). If pressure is now applied, the load will be distributed far more widely over the lower, flatter surface than over the point. Hence the point will flow and spread and, though in so doing its oxide layer may be disrupted and thus expose the metal surface, the distribution of pressure ensured by the other, largely intact, oxide layer reduces the loading intensity on the opposing metal surface. The oxide layers are even more effective when the initial point contacts occur between areas of low curvature. On loading, conditions then approximate to the case of the pressing together of two plane surfaces separated by adhesive layers. Not until extensive plastic flow takes place at both metal surfaces will their oxide films be broken and metallic contact be established. Since such contact cannot take place before the disruption of both oxide layers, it would be impossible to press two plane surfaces into metallic contact where one surface is soft and the other hard, unless the softer metal in flowing plastically were to shear the oxide layer off the hard surface. This occurs when, for example, a lead column is compressed between a pair of anodically polished plane surfaces of polycrystalline copper; after extensive flow, the lead begins to form welds with the copper (Demonstration).

We will now consider the behaviour of the oxide layers under normal sliding conditions, that is, after 'running-in' has prepared the surface for full loading. Where the lubricant boundary layers break down, the oxide layers must come into contact and slide over each other. Both films may then rupture, allowing metallic contact and cold or hot welding; or one film may be ruptured and the other survive; or both films may survive, with or without the help of a re-forming of the lubricant boundary layers. Wherever they are in sliding contact, heat will be liberated in the oxide layers.

The facts concerning the oxide layers which bear on the rôle they play in sliding may be briefly summarized as follows. They consist of extremely thin amorphous films strongly held to the metal (ionic binding). The melting points and hardnesses of some of the oxides and other products of possible interest in sliding surface practice are given in the following Table.

It will be evident from the table that the poor protection afforded by the oxide layers to single-crystal surfaces of copper or aluminium sliding under light loads at a low rate cannot be ascribed to either the refractoriness or hardness of the oxide layer. On the contrary, the oxide layer on the harder iron single-crystal surface gives a considerable measure of protection, although the fusibility and hardness of ferric oxide lie between those of cuprous oxide and aluminium oxide. We have also seen that a thin Beilby layer on one of a pair of copper single-crystal surfaces rubbing together somehow increases the resistance of its oxide layer to rupture.

Experiments with pairs of dissimilar metals lead to similar results. Thus a copper single crystal sliding on aluminium under a light load almost immediately seizes (Demonstration), and the welds break in the body of the aluminium; obviously both oxide layers are injured. On the other hand, when a copper single crystal slides over an iron single crystal under a much heavier load, no seizure occurs (Demonstration). Yet the cuprous oxide film is broken in many places

Compound	M.P. (° c.)	Hardness (Moh)	Melting point metal (° c.)
Fe ₂ O ₃ (haematite)	1565	6½	1535
Fe ₃ O ₄ (magnetite)	1538 (decomposes)	6½	
FeS ₂ (pyrite)	1171	6½	
FeS (troilite)	1193	4½	
FeCl ₂ (laurenite)	670	Soft	
FeOCl	—	—	
Fe ₃ P	1290	—	
Fe ₂ P	1100	—	
(Fe, Ni) ₉ S ₈ (pentlandite)	878	3½-4	
FeCr ₂ O ₄ (chromite)	—	5½	
NiO (bunsenite)	2090	5½	1455
NiS (millerite)	797	3½	
(Ni, Fe)S ₂ (bravoite)	~ 900	5½-6	
CoO	1935	—	1495
CoS (jaipurite)	> 1165	—	
(Co, Ni, Fe)S ₂ (hengleinite)	~ 1000	5	
Cr ₂ O ₃	1990	—	1615
CrN	1500 (decomposes)	—	
CrF ₃	1000	—	
Ag ₂ S (argentite)	825	2½	960
AgCl (cerargyrite)	455	1½	
MnO (manganosite)	1650	5½	1260
Mn ₃ O ₄ (hausmannite)	1705	5½-6	
MnS ₂ (hauerite)	—	4	
Al ₂ O ₃ (corundum)	2050	9	660
MoO ₃	700	Soft	2620
MoS ₂ (molybdenite)	1185	Soft	
SnO ₂ (cassiterite)	1127 (decomposes)	6-7	232
SnO	< 950 (decomposes)	—	
SnS	882	—	
TiO ₂ (rutile)	2130	6-6½	1800
MgO (periclase)	2500-2800	5½-6	651
ZnO (zincite)	1670	4½	419
ZnS (sphalerite)	1049	4	
Cu ₂ O	1235	—	1083
CuO (tenorite)	1026 (decomposes to Cu ₂ O)	3-4	
CuCl (nantokite)	422	—	
Cu ₂ S (chalcocite)	1100	3	
CuS (covellite)	220 (decomposes)	—	
PbO ₂ (plattnerite)	290 (decomposes)	5½	
PbO (litharge)	888	2	327
PbCl ₂ . 2PbO (mendipite)	693	3	
PbO . PbSO ₄ (lanarkite)	977	2½	
PbSO ₄ (anglesite)	1000 (decomposes)	3	
PbFe ₄ O ₇ (plumboferrite)	1530	5	
CdO	1426	3	321
CdS (greenockite)	900 (sublimes)	3½	
ZrO ₂ (zirconia)	> 2950	6½	1900
Sb ₂ O ₃	656	2½	630
Sb ₂ S ₃ (stibnite)	550	2	
Bi ₂ O ₃	820	4½	271
V ₂ O ₃	1970	—	1720
InO ₂	700 (sublimes)	—	155
Spinels (Mg, Fe, Ni, Cr, Mn, or ZnAl ₂ O ₄)	1800	6-8	
Magnetites (Mg, Fe, Zn, or NiFe ₂ O ₄)	1500	5½-6½	
Chromites (Mg or FeCr ₂ O ₄)		5½	

Among other surface compounds of possible interest are the selenides, tellurides, arsenides, phosphates and borates.

and the structure of the underlying metal is disturbed; the iron oxide, however, survives and thus prevents metal-to-metal contact. The oxide layers on surfaces of metals of high rigidity, such as hardened alloy steels, resist rupture at low sliding speeds, even when heavily loaded. Examples are afforded by ball-bearing and gear-tooth surfaces in both of which, although the load intensity is high, the sliding speeds are low or very low. Such bearings, even without a lubricant, are for a time capable of carrying surprisingly heavy loads at low speeds without forming welds. It is evident, therefore, that the resistance of the oxide layer to injury during sliding must, irrespective of its nature, depend in the first place on the rigidity of the supporting metal.

As the severity of the sliding conditions increases, however, it becomes apparent that the behaviour of the oxide layer is determined not only by the supporting metal but also by its own properties. For example, when the bearing surfaces consist of the same hard metal, the frequency of intermetallic contact and welding increases when the sliding speed or load is raised. As we shall see later, however, a change in the chemical composition of the oxide layer alone may reduce or even eliminate intermetallic contact and welding.

It seems to me that these facts are best explained as follows. The ratio of the load to the area of contact is constant and a measure of the rigidity of the softer of the two metals of the opposing surfaces. Hence, with a given load, the loading intensity increases with the rigidity of the softer metal or with that of the metal common to both surfaces. Therefore, since heat is developed in the oxide layers wherever their surfaces are sliding in contact, the local temperature attained must, under comparable conditions, be higher with rigid metals than when one or both are soft. Further, these local temperatures must increase with the speed of sliding. High local temperatures have, indeed, been observed during the sliding of glass or quartz surface pairs (Bowden, Stone and Tudor 1947), and this suggests that the factor limiting the possible temperature rise is the fusion of the oxide layer. If the softening temperature of the oxide layer is higher than that of the metal, the latter will melt and thus no longer support the oxide layer, and the ensuing hot metal-to-metal contact will lead to the formation of a hot weld.

To test this view we may consider the hypoid gear which was introduced about 1930. Though the sliding speeds in this type of gear are low, the load on the bearing surfaces is unusually high, and excessive wear due to welding was common. Some years previously it had been found that wear in cutting tools could be reduced by adding finely divided sulphur to the cutting lubricant. The sulphur seemed to act by forming a protective film on the tool tip. When, therefore, the best of the then known gear lubricants, such as castor oil and metallic soaps, had failed to give proper protection to the hypoid gear surfaces of hard ferrous alloy, it was a short but important step to the introduction by Evans (1937) of oil-soluble organic sulphur compounds which form surface films capable of preventing metal-to-metal contact under severe conditions of both loading and sliding speeds. Since then a bewildering variety of so-called 'extreme-pressure' additives has been increased by compounds containing chlorine or phosphorus. In 1942, Evans showed that the additive molecules did not perform their function by merely becoming adsorbed as such on the bearing surface, but found that a useful working hypothesis was to suppose that sulphide, chloride or phosphide layers were formed. Since then the formation of an unidentified iron sulphide compound, possibly FeS_2 ,

(Simard, Russell and Nelson 1941), of an iron phosphide (Fe_3P) (Beeck, Givens and Williams 1941, Brummage 1942), ferrous chloride (FeCl_2) (Brummage 1942) and ferric oxychloride (FeOCl) (Brummage 1949, private communication) on ferrous bearing surfaces lubricated by oils containing the appropriate additives has been observed. Other layers produced in a similar manner have been identified on α -brass (cuprous chloride, CuCl) (Bryant 1936) and on silver (silver sulphide, Ag_2S). The formation of a sulphur compound layer has also been confirmed by the use of an additive containing radioactive sulphur (Clark, Gallo and Lincoln 1943). Thus Evans' view that the extreme-pressure additive forms a layer by a chemical attack can be accepted. Up to the present, however, no reasonable hypothesis has been put forward to account for the remarkable resistance of this layer to rupture. Indeed, wear of the properly lubricated hypoid gear is so slight that metal-to-metal contact and welding must be of rare occurrence.

Now, it is well known that an injury to the oxide layer on, for example, a ferrous bearing surface is almost immediately healed on renewal of the boundary layer owing to the action on the exposed metal of air and other oxidizing products dissolved in the ordinary lubricant. It is clear, therefore, that the superior protection afforded by the sulphide (or chloride or phosphide) layer formed by the extreme-pressure additive is due to some property inherent in the nature of this surface compound. It cannot be a question of hardness because, excepting pyrite, the sulphides, chlorides and phosphides of iron are softer than the oxides. A significant fact, however, is that, with few exceptions (the noble and platinum-group metals, tantalum, molybdenum and tungsten), the oxides likely to be formed on bearing surfaces are more refractory than the corresponding metals. On the other hand, the sulphides, chlorides and phosphides of the chief constituents of the alloy steels used in hypoid gears are more, or even much more, fusible than the metals. Finally, a zone of sliding contact between the oxide layers is a region of high temperature and, possibly, high pressure.

With these facts in mind, it seems to me that the extreme-pressure additive acts in the following way. Before the frictional temperature at areas of rubbing contact becomes high enough to soften the underlying metal and thus rob the oxide layer of its support, the additive reacts with the oxide to form a layer of a compound, such as a sulphide or oxysulphide, which is more fusible than the metal. If the local temperature should rise sufficiently to melt the compound, some of the molten layer will smear and thus act for a time as a lubricant. The one or two molecular layers next to the metal, being more strongly held, will, however, remain intact and continue to resist rupture, because the melting point of the compound is too low to impair the rigidity of the metal and thus deprive these last molecular layers of their support. This view is further supported by Greenhill's (1948) experiments which showed that the friction between sliding pairs of steel or copper surfaces was much reduced on replacing the oxide layers by sulphides.

It has already been remarked that it is a common experience that in time both bearing surfaces of a sliding element become polished, no matter how widely their melting points or hardnesses may differ. At first sight this seems to conflict with the results obtained by Bowden and Hughes (1937) who found that a soft material could polish a hard one of lower melting point without itself being polished.

The explanation lies in the difference between the conditions under which their experiments were carried out and those of normal sliding-surface practice. There are, indeed, conditions in which hardness and fusibility are of secondary importance. For example, platinum (H.4:M.P.1,775°) is polished by a high-speed buff dressed with rouge (H.6½:M.P.1,565°). A diamond tool tip (H.10:M.P.>3,500°) is worn and smoothed by corundum (H.9:M.P.2,050°) during the turning of aluminium. A rotating pad dressed with corundum polishes pericase (H.5½-6:M.P.2,500-2,800°). Synthetic fibres wear and polish the hard steel surfaces over which they slide in textile machinery. The end of a rod of camphor (M.P.178°c.) is caused to flow and be polished by a swiftly moving ribbon of Wood's metal (M.P.73°c.) (Spurr, recent experiments carried out in this laboratory). These examples suffice to show the importance of the effect of a continual changing or otherwise of the rubbing surface. When two surfaces are in rubbing contact, the heat developed is shared by both. If one surface is always in rubbing contact, e.g. the tool tip, its temperature will exceed that of the other surface which is continually changing, despite conduction losses. For example, a small steel specimen describing a continually changing path on a brass surface can be run without seizure under much more severe conditions than vice versa (Brummage, private communication). In both cases, the small specimen reaches a high temperature, which the steel, because of its higher melting point, is better able to withstand. This is precisely what happens on running in a sliding element pair consisting of a hard and a soft metal. The soft metal is first polished and smoothed. The asperities on the hard surface then form small surfaces which continue to rub in relatively long contact paths over rounded hill tops on the softer surface and are thus in their turn smoothed and polished. In general, where high spots are in sliding contact, the one of smaller radius of curvature will attain the highest temperature, and this may well exceed the melting point of the layer on the surface of larger radius of curvature.

§ 6. THE BOUNDARY LAYER

The function of the oxide layers is to prevent intermetallic friction with its more or less destructive consequences. The oxide-layer (or sulphide etc.) component of friction, however, is still too high to be tolerated in general sliding practice, even supposing the layers could remain intact. It is the function of the boundary layers of the lubricant to replace the oxide-layer component of friction by one of still lower value.

The lubricants consist, in general, of mixtures of straight-chain hydrocarbon molecules (paraffins) or their end-substituted products (alcohols, esters, ketones, acids). At temperatures below the melting point of the hydrocarbons, the chains, 4.5 Å. apart, are grouped in bundles, like sticks in fascines. Such bundles are in fact crystals. When grown from the melt or solution on a solid surface the crystals are orientated with the chains either vertical or inclined to the surface at an angle characteristic of the compound. The thickness of such an orientated crystalline layer may be as much as 50 molecules or more (for references see Brummage 1947).

When such layers are heated, as they are under practical conditions of sliding, the fascines break up at a temperature close to the melting point, leaving a true boundary layer of molecules. The molecules attached directly to the surface

are normal or steeply inclined to it; above them is a layer of less pronounced orientation, and by the time the fourth layer is reached the orientation is practically non-existent (Finch and Zahoorbux 1937); that is to say, the film has now become fluid. The maximum thickness of the boundary layer is therefore of the order of less than 200 Å.

The chains of the first layer in the boundary film are all adsorbed with one end attached to the oxide film, the radicle end in the case of end-substituted molecules. Owing to the polar nature of these, they are more firmly held than the normal paraffins. Further, acids or esters may react with the oxide, sulphide or other layer to form soaps; in this case the relatively thick crystalline layer persists up to the softening temperature of the soap (Brummage 1947, Tingle 1950).

The transition from a crystalline to a boundary film is reversible in that, on cooling, the crystals are re-formed (Brummage 1947). In consequence, new orientations of the crystals or fascines are observed after rubbing. For example, solidified fatty acid films are frequently found to contain crystals in which the molecules are aligned in the direction of sliding at an angle of about 5° to the surface (Finch 1938, Germer and Storks 1939).

Boundary-layer friction is lower with end-substituted than with normal paraffins of similar chain lengths (Hardy 1936). This has generally been attributed to their stronger adhesion to the oxide layer. It is difficult, however, to see how this can materially reduce friction. It may be said, indeed, that present views on the mechanism of the facilitating of sliding by the lubricant boundary layer are inadequate. It is true that I have at times likened the boundary layer to a velvet pile or the bristles of a brush, and have also drawn attention to the thermal motion of the molecules of the layer and to possible effects of their polarity. But all these ideas are at best working hypotheses meant only to serve as guides to further experimental enquiry.

§7. THE INTER-BOUNDARY LUBRICANT FILM

The inter-boundary lubricant film exists only in regions of relatively light loading. Osborne Reynolds (1886) recognized its fluid character, and his theory of hydrodynamic flow not only represents the first broad generalization in the science of lubrication but has also dominated the geometrical design of the sliding element to this day. This is as it should be, for the shear strength of the fluid film is less than that between any of the other surface pairs involved in the sliding element. Boundary conditions of sliding should therefore be restricted as far as possible to those areas of heavier loading where the fluid film can no longer survive as such.

In addition to keeping the boundary layers apart and reducing friction in the more lightly stressed regions, the fluid film has other important functions to fulfil. It must replenish supplies to the boundary layers, bring up oxidizing agents (oxygen, sulphur, halogens etc.) to repair damage sustained by the oxide layers or suitably to modify these, and carry off debris, clothing each particle with a protective boundary layer. Finally, the fluid layer also helps to carry off frictional heat.

The main sources of wear in hydrodynamic lubrication are chemical corrosion and erosion. Erosive wear is chiefly due to debris rolling (Figure 6) between or striking sliding surfaces, and to cavitation (Munroe; see Rayleigh 1917) effects resulting from the collapse of bubbles suspended in the oil and in contact with the sliding surface.

§ 8. CONCLUSION

The problem of the sliding element is a complex one. A distinguished physicist once spoke of it to me as 'a very messy business'. This complexity together with the success of the empirical approach has undoubtedly deterred scientific research into the subject. Since Osborne Reynolds' triumphant framing of the theory of hydrodynamic lubrication, only four schools have attempted to carry the matter further. Beilby's (1921) work on surface flow and Hardy's (1936) views on surface forces were destined, as we have seen, to prove of fundamental importance. The other two schools, both active to-day, are those of Bowden in Cambridge and my own in South Kensington. Hitherto all these workers have been concerned mainly with the metallic bearing surface and the oil *qua* lubricant. If, in this lecture, I have laid particular emphasis on the oxide layer and on the oil as purveyor of oxidant to the high temperature zone of solid-to-solid sliding contact, it is because I believe that it is in this direction that the next great advance will be made. Without the oxide layer and oil-borne oxidants, whose interaction is confined to regions of high temperature, sliding under boundary conditions would be a practical impossibility.

ACKNOWLEDGMENTS

I have to thank Professor E. N. da C. Andrade and Dr. J. W. Fox for Figure 20, M. M. Bluhm for Figure 6, Dr. K. Leu for Figures 16 and 17, D. P. D. Webb for Figure 2, Dr. H. Wilman and D. Layton for Figure 3 and R. T. Spurr for Figures 5, 8 and 9 and for his help, particularly with the single crystal experiments.

My thanks are also due to Dr. K. G. Brummage (Anglo-Iranian Oil Company's Research Laboratories) and to Mr. E. A. Evans (C. C. Wakefield Company's Research Laboratories) for helpful discussions, to Mr. D. Green (Vandervell Limited) for the loan of bearing liners exhibited at my lecture and to Mr. G. Whipple (Hilger and Watts Ltd.) for the loan of a set of slip gauges.

I also wish to record my thanks for financial assistance given to the work carried out in this field in my laboratories by E. G. Acheson Ltd., Anglo-Iranian Oil Company Ltd., Ferranti Ltd., C. C. Wakefield Ltd., the Department of Scientific and Industrial Research, and the University of London Central Research Fund (for the provision of a 200 kv. generator).

REFERENCES

- AMONTONS, G., 1699, see *Encyclopaedia Britannica*, 13th Ed., 1926.
 ANDRADE, E. N. DA C., and FOX, J. W., 1949, *Proc. Phys. Soc. B*, **62**, 483.
 BEECK, O., GIVENS, J. W., and WILLIAMS, E. C., 1941, *Proc. Roy. Soc. A*, **177**, 103.
 BEILBY, G., 1903, *Proc. Roy. Soc. A*, **72**, 218; 1921, *Aggregation and Flow of Solids* (London: Macmillan).
 BOWDEN, F. P., and HUGHES, T. P., 1937, *Proc. Roy. Soc. A*, **160**, 575.
 BOWDEN, F. P., and RIDLER, K. E. W., 1936, *Proc. Roy. Soc. A*, **154**, 640.
 BOWDEN, F. P., STONE, M. A., and TUDOR, G. R., 1947, *Proc. Roy. Soc. A*, **188**, 329.
 BOWDEN, F. P., and TABOR, D., 1939, *Proc. Roy. Soc. A*, **169**, 391; 1945, *Annual Reports Chem. Soc.*, **42**, 20.
 BRUMMAGE, K. G., 1942, *Ph.D. Thesis, London University*; 1947, *Proc. Roy. Soc. A*, **188**, 414, 191, 243.
 BRYANT, F. J., 1936, *Ph.D. Thesis, London University*.
 CLARK, N., and MOTT, N. F., 1949, *Rep. Prog. Phys.*, **12**, 163.
 CLARK, G. L., GALLO, S. G., LINCOLN, B. H., 1943, *J. Appl. Phys.*, **14**, 428.
 COCHRANE, W., 1938, *Proc. Roy. Soc. A*, **166**, 228.

- COULOMB, C. A., 1781, see *Encyclopaedia Britannica*, 13th Ed., 1926, article on 'Friction' by A. B. W. Kennedy and W. E. Dalby.
- DE BROUCKÈRE, L., 1945, *J. Inst. Metals*, **71**, 131.
- DE LA RIVE, A. A., 1839, *Ann. Chim. (Phys.)*, **39**, 328.
- EVANS, E. A., 1937, *Proc. Instn. Mech. Engrs.*, General Discussion on Lubrication; 1942, *Prcc. Instn. Auto. Engrs.*, **37**, 1.
- EWING, A., 1912, *J. Inst. Metals*, **8**, 30.
- FARADAY, M., 1838, see *Experimental Researches in Chemistry and Physics* (London : Taylor and Francis, 1859).
- FINCH, G. I., 1937, *Science Progress*, **31**, 609; 1938, *J. Chem. Soc.*, 1144; 1945, *Proc. Instn. Mech. Engrs.*, **153**, 331.
- FINCH, G. I., and QUARRELL, A. G., 1933, *Proc. Roy. Soc. A*, **141**, 398.
- FINCH, G. I., QUARRELL, A. G., and ROEBUCK, J. S., 1934, *Proc. Roy. Soc. A*, **145**, 676.
- FINCH, G. I., QUARRELL, A. G., and WILMAN, H., 1935, *Trans. Faraday Soc.*, **31**, 28.
- FINCH, G. I., and SUN, C. H., 1936, *Trans. Faraday Soc.*, **32**, 852.
- FINCH, G. I., and WHITMORE, E. J., 1938, *Engineering*, **146**, 91.
- FINCH, G. I., and WILLIAMS, A. L., 1937, *Trans. Faraday Soc.*, **32**, 852.
- FINCH, G. I., and WILMAN, H., 1936, *Proc. Roy. Soc. A*, **155**, 345.
- FINCH, G. I., WILMAN, H., and YANG, L., 1947, *Discussions, Faraday Soc.*, **43A**, 144.
- FINCH, G. I., and ZAHOOBUX, F. D., 1937, *Proc. Instn. Mech. Engrs.*, General Discussion on Lubrication.
- GERMER, L. H., and STORKS, K. H., 1939, *Phys. Rev.*, **55**, 648.
- GREENHILL, E. B., 1948, *J. Inst. Petrol.*, **34**, 659.
- HARDY, W. B., 1936, *Collected Scientific Papers* (Cambridge : University Press).
- HARDY, W. B., and HARDY, J. K., 1919, *Phil. Mag.*, **38**, 32.
- HOPKINS, B. F., and PALM, J. V. O., 1932, British Patent No. 409,289 (accepted 1934).
- HOPKINS, H. G., 1935, *Trans. Faraday Soc.*, **31**, 1095.
- LANGMUIR, I., 1916, *J. Amer. Chem. Soc.*, **38**, 2269; see also, 1920, *Gen. Elect. Rev.*, **23**, 503.
- PARKER, R. C., and HATCH, D., 1950, *Proc. Phys. Soc. B*, **63**, 185.
- PRESTON, G. D., and BIRCUMSHAW, L. L., 1936, *Phil. Mag.*, **22**, 654.
- RAYLEIGH, Lord, 1917, *Phil. Mag.*, **34**, 94.
- REYNOLDS, O., 1886, *Philos. Trans.*, **177**, 156.
- SIMARD, G. L., RUSSELL, H. W., and NELSON, H. R., 1941, *Industr. Engng. Chem.*, **33**, 1352.
- THIRSK, H. R., and WHITMORE, E. J., 1940, *Trans. Faraday Soc.*, **36**, 565, 862.
- TINGLE, E. D., 1940, *Trans. Faraday Soc.*, **46**, 93.
- VERNON, W. H. J., WORMWELL, F., and NURSE, T. J., 1944, *J. Iron and Steel Inst.*, **150**, 81.
- WILLIAMS, C. G., 1937, *Proc. Instn. Mech. Engrs.*, Discussion on Lubrication.
- WILMAN, H., 1950, *Nature, Lond.*, **165**, 321, *Proc. Phys. Soc. A*, **63**, 298.

The Influence of Concentration-Dependent Diffusion on Rate of Evaporation

By J. CRANK

Courtaulds Limited, Maidenhead, Berks.

MS. received 18th January 1950

ABSTRACT. It has been suggested that evaporation through a membrane or from a polymer film has a characteristic behaviour different from that commonly associated with a free liquid surface. In particular it is claimed that an increase in surface humidity produces an increased rate of evaporation through a membrane, though the opposite effect is well known for a liquid surface. This behaviour is said to be observed when diffusion within the membrane or film is the rate controlling process in systems for which the diffusion coefficient increases with increasing concentration.

In this paper the behaviour of such diffusion-controlled evaporation processes is examined from the point of view of the diffusion equations and it is found that there is no theoretical reason to expect such processes to show anomalous behaviour.

It is seen that when the experimental data for evaporation through a membrane are correctly interpreted they are in agreement with theoretical prediction.

The so-called 'skin effect' in the dry spinning of fibres is examined and appears to be due to complicating features other than diffusion.

§ 1. INTRODUCTION

MANY evaporative processes are diffusion-controlled in the sense that the rate of evaporation depends largely on the rate at which solvent or moisture is supplied to the evaporating surface by internal diffusion. The dry spinning of fibres and the loss of moisture from the body through the skin are examples of particular interest to the textile industry.

It is of considerable practical importance to know how the rate of evaporation is influenced by external conditions, particularly those over which control can be exercised. Thus in the evaporative (dry) spinning process, for example, not only the rate of production of fibres but also the shape of the filaments and their properties are profoundly affected by the rate of evaporation of solvent at different stages of the process.

The present paper is concerned with the effect of the proportion of solvent vapour, or the relative humidity, in the atmosphere into which the evaporation takes place. In the case of evaporation from a free liquid surface it is well known that the higher the vapour pressure in the external atmosphere the less the rate of evaporation and vice versa. Correspondingly, the higher the wind speed past the exposed surface, the greater the rate of evaporation because the thickness of the boundary layer through which the vapour must pass in evaporating decreases with increasing wind speed. With this in mind it is somewhat surprising to find that several instances of the opposite behaviour have been reported, where an increase in relative humidity at the evaporating surface produces an increased rate of evaporation. This is said to happen when evaporation occurs through an organic membrane or from a polymer film, in which cases the supply of solvent to the surface is governed by diffusion. The following are examples of this anomalous behaviour.

Throughout the patent literature of dry spinning (see for example British Patent Specification 238,842) there is the suggestion that the rate of loss of solvent from the filaments, e.g. of acetone from cellulose acetate filaments, can be increased by including a large proportion of solvent vapour in the atmosphere through which the filaments pass after issuing from the dies. This behaviour is attributed to the fact that when the evaporation takes place into an atmosphere containing little or no solvent vapour a dry surface layer, frequently called a skin, is formed almost as soon as the filaments leave the dies. As it is known that the diffusion coefficient is very small in a near-dry polymer, the skin is considered to act as a barrier to the subsequent diffusion of solvent from the filament core and hence to slow down the evaporation in the later stages. The formation of the skin is prevented by using a drying atmosphere rich in solvent vapour, and it is therefore argued that a more rapid removal of solvent from the filaments is obtained.

More recently Cassie (1949), discussing the characteristics for warmth in underwear fabrics, has stated that, at low temperatures, the heat loss by evaporation from the body increases as the relative humidity surrounding the skin increases. This he considers to be a well-known effect for membranes such as the skin and in support of the observation refers to the diffusion coefficient for water through a keratin membrane, the coefficient being relatively very small at low moisture contents.

The purpose of the present paper is to show mathematically that, if such anomalous behaviour exists, it cannot be attributed simply to the low value of the diffusion coefficient in the near-dry film or membrane. Calculations show that any abnormality in behaviour must be due to complicating features other than the simple diffusion process. In order to reduce the mathematical complications to a minimum, the main calculations are carried out for a lamina instead of for a cylinder. Effects such as puckering of the surface of the filaments are neglected.

Part I of the paper deals with the steady-state flow through a membrane, and Part II with the rate of the loss of solvent from a polymer film.

PART I

§2. STEADY-STATE EVAPORATION THROUGH A MEMBRANE

(i) *Theoretical*

If we have a membrane of thickness l separating a region of high relative humidity from one of low relative humidity, then, provided the humidities on the two sides are maintained constant, a steady-state transfer of moisture through the membrane is set up. The concentrations C_0 and C_1 just within the two surfaces of the membrane are constant and are the equilibrium regains for the high and low humidities respectively. A concentration gradient exists through the membrane, and moisture moves down the gradient by diffusion at a rate determined by the concentrations C_0 and C_1 and the diffusion coefficient D . The rate of evaporation F per unit area of the membrane is given by

$$F = -D \frac{dC}{dx}, \quad \dots\dots(1)$$

which on integration becomes

$$\int_0^l F \, dx = \int_{C_1}^{C_0} D \, dC. \quad \dots\dots(2)$$

Since we are concerned with a steady state, F is the same through each section of the membrane, i.e. F is independent of x . Therefore we have

$$\int_0^l F dx = F \int_0^l dx = Fl, \quad \dots\dots(3)$$

and finally

$$F = \frac{1}{l} \int_{C_1}^{C_0} D dC. \quad \dots\dots(4)$$

The above argument is true whether D is constant or not. In particular it is true when D is a function of concentration, and provided D is never negative in the range from C_0 to C_1 , the integral in (4) must always increase or remain constant as the range of concentration is increased. Thus if the high concentration C_0 remains fixed, the rate of flow F can never decrease as a result of decreasing the lower concentration C_1 . The significance of (4) in this connection was brought to my notice by my colleague, Dr. G. S. Park.

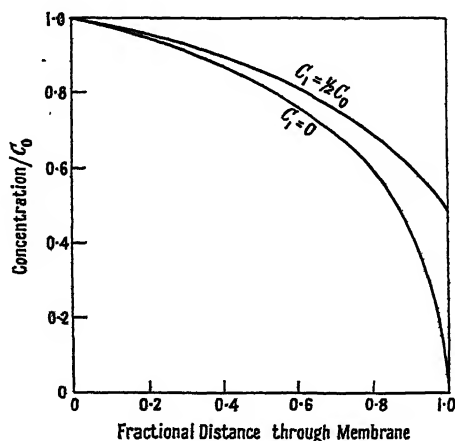


Figure 1. Concentration distribution through a membrane for a diffusion coefficient increasing exponentially with concentration. C_0 , C_1 are the concentrations on the high and low pressure sides respectively.

Figure 1 shows calculated concentration distributions through the membrane when D is an exponential function of C such that D increases by 50-fold from $C=0$ to $C=C_0$. The concentration C_0 on the high humidity side is the same for each curve but in one case $C_1=0$ and in the other $C_1=\frac{1}{2}C_0$. The curves show how the concentration distribution adjusts itself so that the rate of flow can be greater when $C_0=0$ even though the region of low diffusion coefficient is then included. Clearly a low diffusion coefficient is compensated by a high concentration gradient.

(ii) Experimental

In a recent paper Cassie (1949) remarks that it is well known that an increase in relative humidity on the low humidity side of a membrane results in an increased rate of evaporation. He describes a convenient way of demonstrating the effect experimentally, by observing the rate of evaporation from a collodion bag filled with water and reports that "the evaporative loss from the bag actually decreases as the wind speed increases because the higher wind speed maintains

a low relative humidity at the surface of the bag". The behaviour described appears plausible on the grounds of the intuitive reasoning outlined earlier in the paper, i.e. that increasing the surface humidity prevents the formation of a near-dry surface layer through which diffusion is slow. Thus there appears to be a discrepancy between the behaviour predicted by equation (4) on the one hand, and that expected intuitively and observed experimentally on the other.

The original observations were made by using the collodion bag as a wet bulb thermometer. An increased wind speed was accompanied by an increase in temperature and from this it was concluded that the rate of evaporation had decreased. The experiment has recently been repeated, however (Cassie, private communication), and, whilst the temperature change is confirmed, a direct measurement of the rate of evaporation reveals that this does in fact increase as the wind speed increases. This is in agreement with the mathematical prediction of equation (4) above. If a temperature increase accompanies the increase of wind speed (probably due to an increase in heat exchange from the moving air) it follows from (4) that the observed increase in evaporation will be greater than if the temperature remained constant, because the diffusion coefficient D increases with increasing temperature. Thus we conclude that the rates of evaporation through a membrane and from a free liquid surface are both decreased by increasing the surface humidity or decreasing the wind speed.

PART II

§ 3. EVAPORATION OF SOLVENT FROM POLYMER FILMS AND FILAMENTS

Physical Model on which Calculations are based

It is assumed that the loss of solvent from the polymer film takes place by diffusion through the film with subsequent evaporation from the surface exposed to the drying air stream. On this assumption the course of drying may be calculated from the diffusion equation. We have seen that the diffusion coefficient is generally concentration dependent in such systems, and the nature of this variation and also the law governing the surface evaporation must be known before the calculations can proceed.

The calculated results shown below relate to the loss of acetone from a plane sheet of cellulose acetate exposed to a drying atmosphere on one surface only. The concentration of acetone is initially uniform throughout the sheet and constant temperature conditions are assumed. Two kinds of condition at the exposed surface are considered:

(i) The surface of the sheet is assumed to reach equilibrium with the drying atmosphere instantaneously when drying commences, i.e. if the atmosphere is acetone-free, the concentration at the surface falls instantaneously to zero; if the vapour pressure in the atmosphere here is p , the surface of the sheet immediately reaches the concentration which is in equilibrium with p .

(ii) A condition expressing the rate of evaporation of acetone from the surface is assumed. The simplest reasonable condition is that the rate of evaporation is directly proportional to the difference between the actual concentration, C_s , in the surface of the sheet at any time and the concentration C_0 , which would be in equilibrium with the vapour pressure in the atmosphere remote from the surface. Mathematically this means that at the surface

$$D_s \partial C / \partial x = \alpha (C_s - C_0), \quad \dots\dots(5)$$

where α is the constant of proportionality and D_s is the value of the diffusion coefficient when the concentration is C_s ; x is the space coordinate perpendicular to the face of the sheet, increasing positively inwards.

The calculated rates of loss of solvent given below are obtained by evaluating solutions of the diffusion equation

$$\frac{\partial C}{\partial t} = \frac{\partial}{\partial x} \left(D \frac{\partial C}{\partial x} \right), \quad \dots\dots(6)$$

with a surface condition of the type (i) or (ii) above. Only the results of the calculations are given here. Details of methods of evaluating numerical solutions of the diffusion equation with a variable diffusion coefficient, and taking account of any shrinkage of the film accompanying loss of solvent, are described elsewhere (Crank and Henry 1949).

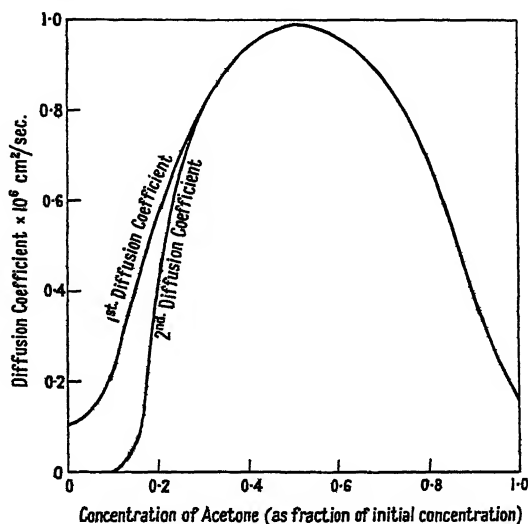


Figure 2.

The first diffusion coefficient plotted in Figure 2 is derived from observations of the rate of penetration of acetone into commercial cellulose acetate sheet, using an interferometer technique described by Crank and Robinson (1950). The significance of the second diffusion coefficient shown in Figure 2 is discussed later.

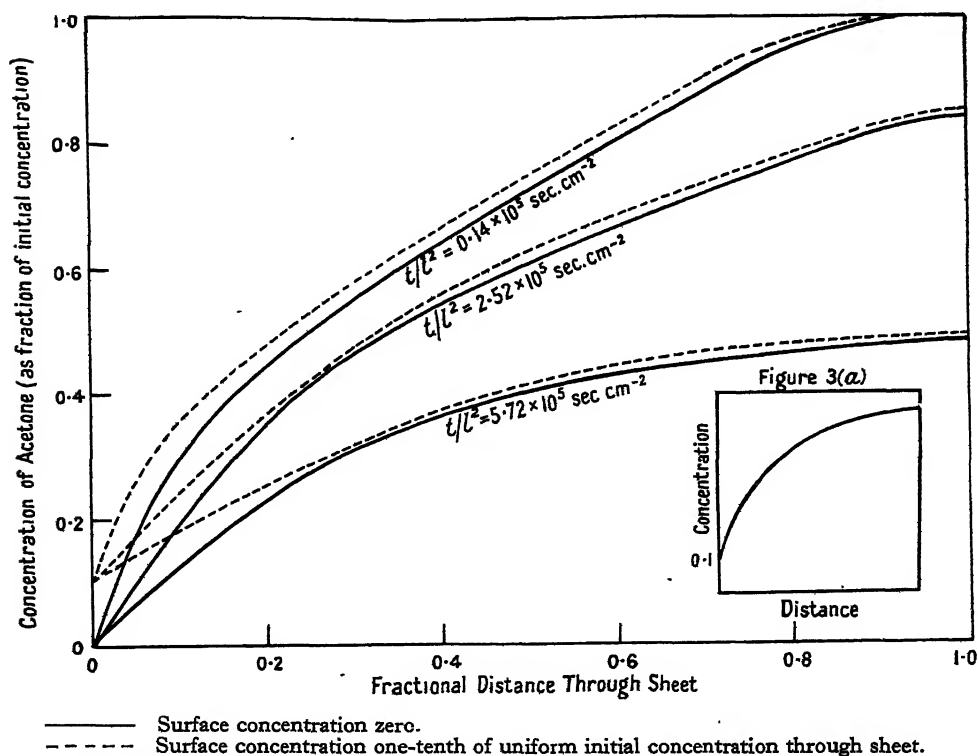
§ 4. CALCULATED RATES OF DRYING AND DISTRIBUTION OF ACETONE

(i) *Surface taking up its final concentration instantaneously.*

This is the condition of vigorous surface evaporation. Figure 3 shows the calculated variation of concentration of acetone with distance through the sheet at three successive times for two different surface concentrations; (1) the surface concentration falls to zero instantaneously, (2) it falls to one-tenth of its initial value instantaneously. There is little difference between the two curves at any time. In particular the total acetone content of the sheet at any time, represented by the area under the appropriate curve of Figure 3, is much the same whether the surface concentration falls to zero or to one-tenth its initial value, i.e. whether

a concentration range over which D is small is removed or not. In so far as there is a slight difference, the drier sheet at any time is obtained by putting the surface to zero concentration, which means vigorous evaporation into an acetone-free atmosphere. No advantage is gained by introducing acetone into the atmosphere.

The reason for this is clear from the calculations, as it was in the steady-state case shown previously in Figure 1. When the value of D at the surface is very low the concentration gradient is correspondingly large so that the overall rate of loss of acetone, given by the product $D\partial C/\partial x$ at the surface, does not alter appreciably. This compensating effect is most obvious when the diffusion coefficient is zero over a range of low concentrations as in the second diffusion coefficient shown in Figure 2. This is a hypothetical coefficient chosen to exaggerate the effect. The concentration-distance curve for this coefficient is sketched in Figure 3(a) from general reasoning. The surface gradient is infinite over the concentration range for which D is zero (in this case for concentrations



less than 0.10) so that the product can have a non-zero value. A finite gradient cannot develop for concentrations less than 0.10, since this would imply that solvent had been removed from a region of zero diffusion coefficient under the action of a finite gradient, which is not possible.

Thus if we interpret the term 'skin' to mean a gel whose acetone content is less than a certain critical amount, so that the diffusion coefficient in the skin is low, we find that the lower the mean diffusion coefficient in the skin the greater the concentration gradient throughout the skin which is therefore thinner. With this kind of surface condition, therefore, it appears that it is not possible

to form, by diffusion of acetone, a skin which can seriously impede further diffusion of acetone.

(ii) *A rate of surface evaporation proportional to $C_s - C_0$.*

Figure 4 shows the concentration-distance curves calculated for the surface rate of loss condition (5) and the first diffusion coefficient of Figure 2. The curves of Figure 4 are for $\alpha l = 16 \times 10^{-6}$; in one case $C_0 = 0$ and in the other $C_0 = 0.2$, measured as a fraction of the initial concentration in the sheet. At any given time the surface is always drier for $C_0 = 0$, but there is the rather interesting feature that the concentration-distance curves cross within the sheet. Thus there is an interval of time during which the inside of the film is drier at a given time when the evaporating atmosphere contains acetone than when it is acetone-free. The corresponding curves for $\alpha l = 8 \times 10^{-6}$ were calculated, and found not to intersect. They are not sufficiently interesting

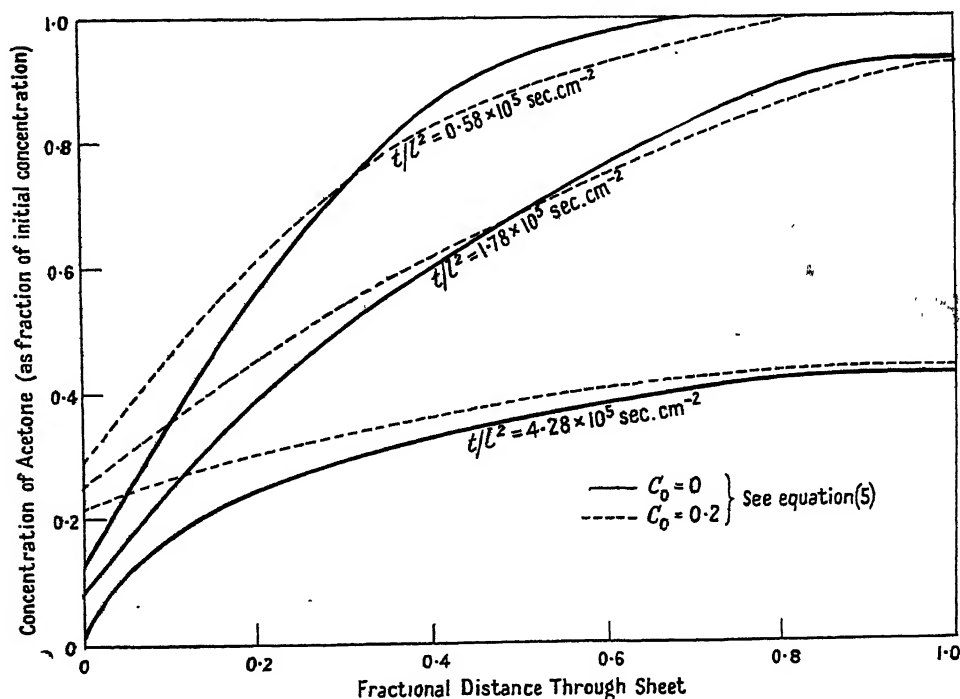


Figure 4. Concentration distributions for first diffusion coefficient of Figure 2, $\alpha l = 16 \times 10^{-6}$.

to show. Since the curves for vigorous evaporative conditions, i.e. $\alpha = \infty$, do not cross either, it is clear that there is a limited range of values of αl for which crossing is observed. Although the magnitude of this effect will vary slightly with different values of αl and C_0 , and different diffusion coefficients, other calculations, detailed results of which are not shown, suggest that the effect will never be very large. Thus for the second diffusion coefficient in Figure 2 the separation between the concentration-distance curve is not appreciably different from that of the corresponding curves in Figure 4.

It is clear from Figure 4, as from Figure 3, that the overall rate of loss of solvent from the film is not significantly increased by introducing acetone into the atmosphere.

§ 5. CYLINDRICAL FILAMENTS

Similar calculations for the rate of loss of solvent from a cylinder show the same general behaviour as for the plane sheet. The most rapid overall loss of solvent is obtained when the drying atmosphere is solvent-free. The crossing of the concentration-distance curves is even less marked than in Figure 4.

§ 6. CONCLUSION

Theoretical considerations of diffusion-controlled evaporation do not appear to support the view that the rate of evaporation either through a membrane or from a polymer film can be increased by increasing the surface humidity. In the case of the membrane, the discrepancy between theory and experiment seems to have been due to a misinterpretation of the observational data and this discrepancy has now been removed.

If the presence of solvent vapour in the atmosphere in other cases really does result in an increased rate of evaporation as reported, it appears that this must be due to factors other than simple diffusion. In the case of the dry spinning process, for example, the tension to which the filaments are usually subjected whilst losing solvent may be an important complicating feature. The degree of stretch, and hence of radial contraction, produced by a given tension must depend very much on whether a near-dry skin is formed or not in the early stages of drying. Thus the main effect of solvent vapour in the atmosphere may be to allow the filament cross section to be decreased considerably by longitudinal stretching. The rate of drying is thereby increased since it is inversely proportional to the square of the radius. However, a detailed investigation of the cause of any apparently anomalous behaviour is outside the scope of the present paper. In all probability the explanation will differ from one system to another.

ACKNOWLEDGMENTS

It is a pleasure to thank Dr. A. B. D. Cassie for a most helpful correspondence concerning the observations on evaporation through membranes, and for his willingness to allow the essential points to be communicated in Part I of the paper. The author is indebted to Miss M. E. Henry who carried out the calculations.

REFERENCES

- BRITISH PATENT SPECIFICATION No. 238,842, Société pour la fabrication de la soie rhodiasata.
CASSIE, A. B. D., 1949, *J. Text. Inst.*, **40**, 444.
CRANK, J., and HENRY, M. E., 1949, *Trans. Faraday Soc.*, **45**, 636.
CRANK, J., and ROBINSON, C., 1950, in course of preparation.

The E.M.F.-Temperature Calibration of Platinum, 10% Rhodium-Platinum and Platinum, 13% Rhodium-Platinum Thermocouples over the Temperature Range 0°-1,760 °C.

By C. R. BARBER

Physics Division, National Physical Laboratory, Teddington, Middlesex

Communicated by Ezer Griffiths; MS. received 31st October 1949, and in amended form 16th January 1950

ABSTRACT. E.M.F.-temperature reference tables are given for platinum, 10% rhodium-platinum and platinum, 13% rhodium-platinum thermocouples over the range from 0° to 1,760° C. The tables are based on the calibration of twelve thermocouples of each kind, six from each of two manufacturers. Each set of six thermocouples comprised two samples from each of three batches of wire. The calibrations are in terms of the International Temperature Scale as revised at the Ninth General Conference of Weights and Measures held in Paris in October 1948.

The methods of intercomparison between the thermocouples and the various interpolation instruments specified in the International Temperature Scale are described. It is shown that the difference between the new tables and the old are mainly due to changes in the International Temperature Scale, and that the new tables represent the shape of the E.M.F.-temperature characteristics of the thermocouples of either make sufficiently closely to give interpolation between calibration points with accuracy within $\pm 1^\circ\text{C.}$ up to 1,063° C., $\pm 2^\circ\text{C.}$ up to 1,550° C., and $\pm 3^\circ\text{C.}$ above this temperature.

§ 1. INTRODUCTION

REFERENCE tables for platinum, rhodium-platinum thermocouples of British origin have hitherto been prepared by the manufacturers of the thermocouples from skeleton tables given by the National Physical Laboratory in reports on the routine calibrations of sample thermocouples over a number of years, and from a reference table for 13% couples published in 1943 (Barber 1943) for the temperature range from 1,400° to 1,750° C. This range of temperature is of importance in steel melting, and the reference table was prepared to meet the needs of the then recently developed quick-immersion technique of liquid steel temperature measurement. Since the International Temperature Scale above 1,063° C. is realized by use of the optical pyrometer, the accurate determination of the form of the thermocouple calibration curve involves a comparison between thermocouple and optical pyrometer. In 1942 other urgent work made it impossible to do this, and less direct methods had to be relied upon.

A reference table of the kind here considered sets out to give the shape of the E.M.F.-temperature curve as precisely as possible. It may be used to interpolate accurately between a small number of calibration points on any particular thermocouple, in any desired temperature range, by plotting the difference of the calibration points from the values in the reference table. The shape of the E.M.F.-temperature curve varies very slightly from one thermocouple to another, presumably on account of minute traces of impurity in the thermocouple wires. The accuracy of the reference table in representing the curvature of the E.M.F.-temperature relation of any particular thermocouple is thus limited,

and the greater these variations are the more calibration points are needed to compose a difference curve which will give an accurate calibration of an individual couple.

The tables given at the end of this paper have superseded those formerly issued by the makers, from which they differ quite markedly above the gold point ($1,063^{\circ}$ C.). This difference is mainly accounted for by the changes in numerical values of temperature in certain ranges resulting from the introduction, on 1st January 1949, of the International Temperature Scale (Comité internationale des Poids et Mesures, 1948), revising the original 1927 scale.

§2. SCOPE OF THE INVESTIGATION

The calibration of the thermocouples in terms of the International Temperature Scale involves a comparison ultimately with the interpolation instruments specified in the scale. For the range from 0° to 630° C. the platinum resistance thermometer is employed, using a quadratic equation,

$$R_t = R_0(1 + At + Bt^2),$$

relating resistance to temperature, based on the melting point of ice (0° C.), the boiling point of water (100° C.), and the boiling point of sulphur (444.60° C.).

From 630° to $1,063^{\circ}$ C. the interpolation instrument is the platinum, 10% rhodium–platinum thermocouple, employing a quadratic equation,

$$e = a + bt + ct^2,$$

relating E.M.F. to temperature, based on the freezing points of antimony (630.5° C.), silver (960.8° C.) and gold ($1,063^{\circ}$ C.).

Above $1,063^{\circ}$ C. the optical pyrometer is used and is calibrated on the basis of the freezing point of gold and the Planck law of radiation with a value of 1.438 for the constant C_2 . A temperature t is then defined by the formula

$$\frac{J_t}{J_{Au}} = \frac{\exp \{C_2/\lambda(T_0 + t)\} - 1}{\exp \{C_2/\lambda(T_0 + t_{Au})\} - 1},$$

where J_t and J_{Au} are respectively the spectral energies radiated at wavelength λ by unit area of a black body in unit time at the temperature t and at the freezing point of gold, t_{Au} . T_0 is the melting point of ice in degrees K. defined for the purpose as 273.15.

The most significant differences in the scale outlined above compared with the 1927 scale are that the silver point is defined as 960.8° C. instead of 960.5° C., the value of C_2 is defined as 1.438 instead of 1.432, and the Planck Law of radiation is used instead of the Wien law. The effect of these differences on the numerical values of temperature is shown in Figure 1, where it is seen that in the region between 630° and $1,063^{\circ}$ C. the temperatures as defined by the 1948 scale are higher than those of the 1927 scale, with a maximum difference of 0.4° C. at about 850° C., and above $1,063^{\circ}$ C. they are lower, progressing from 0° at $1,063^{\circ}$ C. to 1.6° at $1,400^{\circ}$ C. and 3.6° at $1,700^{\circ}$ C. These changes account for the larger part of the departures of the E.M.F. values given in the present reference tables from those formerly supplied by the makers. The adoption of the absolute electrical units in place of the international units on 1st January 1948 caused a rise in numerical values of voltage of 34 parts in 10^5 , and this change must also be taken into consideration when the new tables are compared with the old.

§ 3. CALIBRATION OVER THE RANGE 0° TO 1,063° c.

The determination of the mean E.M.F.-temperature curves for both the 10% and 13% rhodium-platinum types of thermocouple has been based on the calibrations of twelve thermocouples of each kind, six being supplied by Messrs. Johnson, Matthey & Co. Ltd. and six by the Sheffield Smelting Co. Ltd. Each set of six couples comprised two from each of three batches of wire. They were annealed electrically at a temperature of about 1,600° c. and then calibrated with cold junctions in ice by direct comparison with platinum resistance thermometers over the range from 0° to 630° c. in stirred water, oil and salt baths (Grace and Hall 1943).

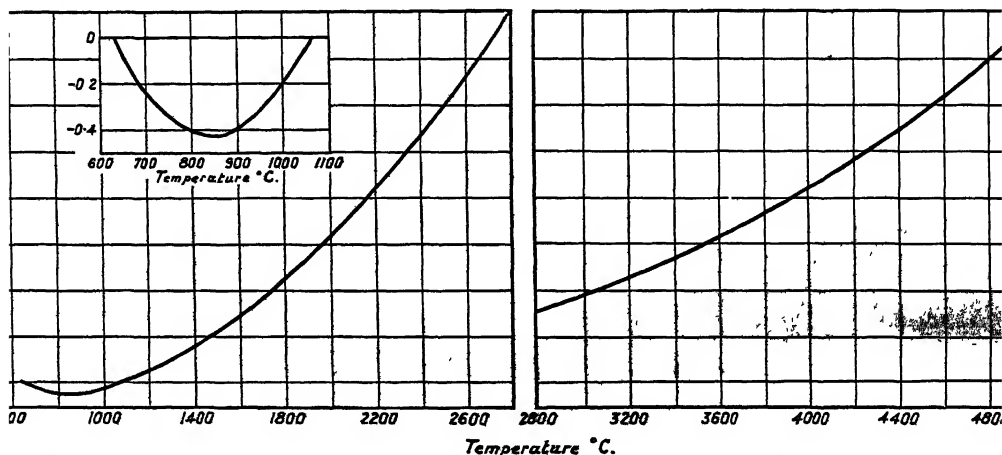


Figure 1. Changes in numerical value of temperature resulting from change from International Temperature Scale 1927 to International Temperature Scale 1948.

The measurements at the freezing point of silver were made on an ingot of pure silver of about 380 gm. weight, cylindrical in shape, of 2.5 cm. diameter and 10 cm. length. The ingot was contained in a graphite crucible and the surface of the ingot was covered with graphite powder to a depth of about 1 cm. The thermocouple was sheathed with a silica tube of 6 mm. external diameter and was immersed coaxially in the ingot to a depth of 9 cm., the silica sheath being protected from attack by the silver by a thin graphite sheath. Uniform temperature conditions over the length of the ingot were obtained by use of a substantial block of graphite for the crucible, which was formed by boring an axial hole 2.5 cm. in diameter in a cylinder about 7.7 cm. diameter and 20 cm. long. A series of freezes was observed on two of the thermocouples to establish the validity of the point. The reproducibility of E.M.F. of the thermocouples from one freeze to another was within 1 microvolt, equivalent to better than 0.1° c. An upward displacement of the thermocouple in the ingot of 2 cm. during a freeze produced no greater change than 0.2 microvolt. Having thus established that satisfactory readings were being obtained, a freezing point observation was made on each of the 24 thermocouples in turn.

For the gold-point observations, an ingot of 570 gm. weight, 2.5 cm. diameter and 6 cm. length was used. Since it is not necessary to protect the gold from the air as with silver, the graphite crucible was dispensed with and a silica crucible used with the thermocouple in a silica sheath immersed directly into the gold.

The crucible was placed in an axial hole bored in a nickel block in order to promote temperature uniformity. A similar degree of reproducibility was obtained as with the silver freeze.

§4. CALIBRATIONS FROM $1,063^{\circ}$ TO $1,760^{\circ}$ C.

The calibrations above the gold point involve comparison either directly or indirectly with the optical pyrometer. The indirect method is to use the freezing points of pure metals already determined on the optical pyrometer scale. The only points easily realized in the temperature range in question are the freezing points of palladium ($1,552^{\circ}$ C.) and platinum ($1,769^{\circ}$ C.). Other metals, such as nickel (freezing point $1,453^{\circ}$ C.) and cobalt (freezing point $1,492^{\circ}$ C.) must be melted *in vacuo* or in a reducing atmosphere. Even if all four of these points were used, however, a large temperature interval with no calibration point would be left between $1,063^{\circ}$ C. and $1,453^{\circ}$ C. The direct method was therefore adopted, and this required a black-body enclosure containing the hot junction of one or more thermocouples, the E.M.F. of which could be measured at the same time as the temperature of the enclosure was determined by means of the optical pyrometer.

A diagram of the furnace used is shown in Figure 2. The black-body enclosure was formed by a hollow platinum cylinder 2 cm. in diameter and wall thickness 1 mm. The back wall of the cylinder was set at an angle to the front wall so that the optical pyrometer should not receive a reflected image of the sighting aperture, which was a circular hole of 2 mm. diameter in the centre of the front wall. The furnace winding consisted of 10% rhodium-platinum foil 12.5 mm. wide and 0.025 mm. thick.

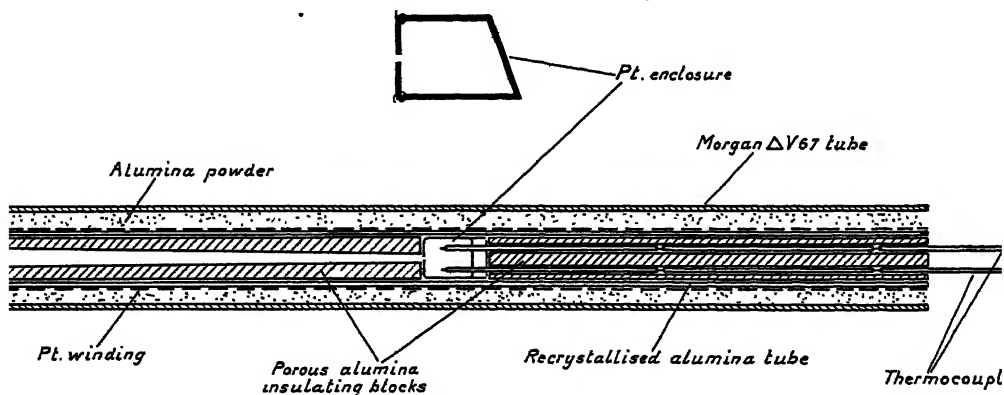


Figure 2. Central furnace tubes and black-body enclosure.

The optical pyrometer employed was the primary standard instrument in which the telescope is divided to permit the use of a rotating sector disc between objective and lamp. It was housed in a constant temperature room and was calibrated immediately before these measurements at the fundamental point of the scale, the gold point, using the furnace arrangement shown in Figure 3. The furnace tube, 2 cm. in bore, was unlagged except for a single radiation shield. The gold ingot, weighing about 30 gm., was contained in a steatite crucible and had an alumina sighting tube of 2.5 mm. bore passing through the axis. In the centre of this tube was a steatite plug through which was drilled a very fine hole.

The furnace was used in the horizontal position and, since it was symmetrical, could be viewed from either end. The more usual form of black-body tube for this purpose is closed at one end and is immersed vertically in an ingot of the metal. It is uncertain with this method that the pyrometer is sighted along the axis of the tube since there is no visible detail in the furnace when uniform temperature conditions exist. The difficulty is overcome in the present arrangement, for the hole in the plug appears as a black spot in the field of the pyrometer when the alignment is correct. As there is only a small amount of furnace lagging, the end-cooling is small in comparison with radiation loss and, since the ingot is relatively short, very uniform temperature conditions are obtained. This is further aided by the fact that the thermal inertia of the furnace is small, and it is therefore only necessary to change the temperature of the windings by

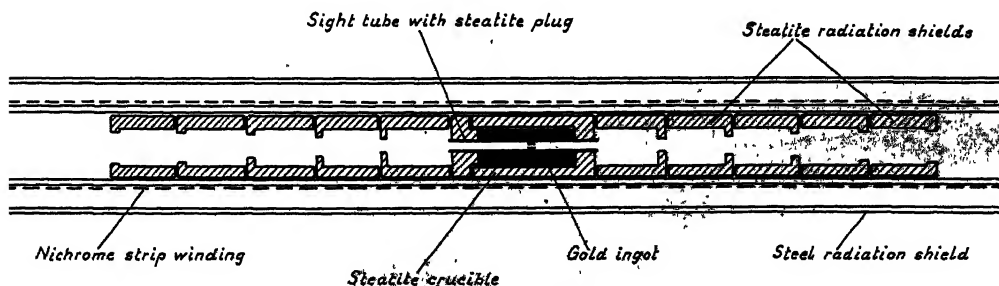


Figure 3. Furnace and ingot assembly for gold point.

a few degrees in order to cause the metal to melt or freeze, so that the temperature gradients are limited. Under these conditions well-defined melting- and freezing-point curves were obtained and no significant differences were detected between observations made by sighting into opposite ends of the furnace.

The field of an optical pyrometer sighted into such a furnace is not uniformly bright. When the ingot is freezing the black-body tube appears as a bright central spot surrounded by darker zones and, when it is melting, as a dark spot with brighter surroundings. Under these conditions the matching of the filament against the central spot is influenced by the differences in intensity of the adjacent parts of the field, and in this determination led to an apparent difference between melts and freezes of 0.7°C . By using a supplementary lens to give a real image of the furnace, and at this position placing a diaphragm of slightly smaller diameter than the image of the black body, the field appears as a central bright spot in complete blackness when the pyrometer is focused on to the diaphragm; agreement between melts and freezes is then obtained.

The pyrometer was then sighted into the platinum enclosure, into which had been inserted thermocouples calibrated at the freezing point of gold as already described. When the furnace was adjusted to give the gold-point temperature as indicated by the E.M.F. of the thermocouples, the optical pyrometer readings agreed, to within 0.3°C ., with those obtained by the direct method, thus giving a check on the efficiency of the black-body radiator. A further check was obtained in a similar manner at the freezing point of palladium ($1,552^{\circ}\text{C}$.), the thermocouples having been calibrated at that point by the wire method and the pyrometer by use of the gold point and a sectored disc of appropriate aperture. The mean difference between the two methods in this case was only about 0.1°C .

Selected thermocouples were calibrated against the optical pyrometer in pairs at temperatures of about $1,200^{\circ}$, $1,270^{\circ}$, $1,310^{\circ}$, $1,420^{\circ}$, $1,550^{\circ}$ and $1,650^{\circ}$ C., each of these temperatures, with the exception of the last-mentioned, being determined by use of a sector'd disc of appropriate aperture to cut down the spectral brightness of the black body, at the temperature named, to that at the gold point. The reduction from $1,650^{\circ}$ C. was achieved in two stages, $1,650^{\circ}$ – $1,200^{\circ}$ and $1,200^{\circ}$ – $1,063^{\circ}$ C. The final measurement at the platinum point by the wire method, in which the platinum arm of the thermocouple is melted at the hot junction, was carried out in a manner similar to that previously described (Barber 1943). The couples calibrated in this way were compared with the remainder, after each point, in a separate furnace held at the same temperature. At the same time, as a check for deterioration in E.M.F., comparison was made with a control couple at the gold point and no significant changes were detected. On completion of all the observations up to and including that at $1,650^{\circ}$ C., the E.M.F. of the thermocouples at the gold point had not changed by more than two microvolts from the original values.

§ 5. DISCUSSION OF RESULTS

Figure 4 shows the departures in E.M.F. of the mean of the six thermocouples of each make from the mean of all twelve thermocouples of one kind. These departures nowhere exceed $5\text{ }\mu\text{V.}$ for either the 10% or the 13% types, and common reference tables may therefore be computed without significant loss in accuracy in representing the true shapes of the E.M.F.–temperature curves. The departures in E.M.F. of individual couples of the 13% type of each make from their own means are shown in Figures 5 and 6. The 10% couples showed similar deviations. Since the wires were not specially selected, the deviations may be taken as representative of those likely to be obtained with any other couples from the same sources.

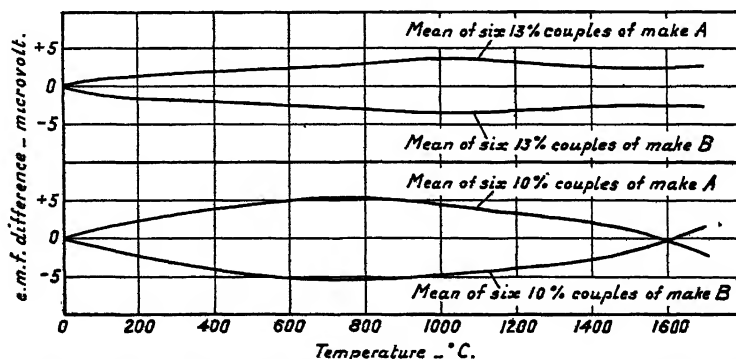


Figure 4. Departures of mean of E.M.F. of six couples of each make from mean of all twelve.

The means of the observations for the twelve 10% and for the twelve 13% couples are recorded in Table 1, and from these values the reference Tables 2 and 3 have been computed by the method described in the Appendix. The difference between the E.M.F. values given by the reference tables and the observational values are shown in columns 4 and 6 of Table 1. Apart from the values at $1,270^{\circ}$ and $1,309^{\circ}$ C. on the 10% couples and $1,270^{\circ}$ C. on the 13% couples, there is no departure greater than $3\text{ }\mu\text{V.}$ (0.2° C.).

The accuracy of the measurements is estimated to be to within $\pm 0.1^{\circ}\text{C.}$ up to $1,063^{\circ}\text{C.}$, $\pm 0.2^{\circ}\text{C.}$ at the gold point as measured with the optical pyrometer, rising to $\pm 2^{\circ}\text{C.}$ at $1,650^{\circ}\text{C.}$ The calibration of a particular thermocouple obtained by measurement at a few temperatures and the use of a difference curve from the reference table will include these inaccuracies and any occurring in the calibration points. On the basis of these considerations it is estimated that observations at the gold and palladium points by the wire method and the use of the tables should yield an accuracy within $\pm 1^{\circ}\text{C.}$ up to $1,063^{\circ}\text{C.}$, $\pm 2^{\circ}\text{C.}$ up to $1,550^{\circ}\text{C.}$ and $\pm 3^{\circ}\text{C.}$ above.

Table 1

Mean Observed Values of E.M.F. of twelve 13 % and twelve 10% Thermocouples

Comparison instrument	Temp. ($^{\circ}\text{C.}$) (Int. 1948)	13% couples		10% couples	
		E.M.F. ($\mu\text{V.}$)	Diff.* ($^{\circ}\text{C.}$)	E.M.F. ($\mu\text{V.}$)	Diff.* ($^{\circ}\text{C.}$)
Platinum wire	1769.0	21107	0.0	18692	-0.2
	1648.5	19514	0.0	17334	0.0
	1552.0	18193	0.0	16204	-0.1
Optical pyrometer	1420.9	16339	-0.1	14622	+0.2
	1308.8	14745	-0.3	13270	-0.6
	1270.3	14192	+0.6	12787	+0.6
	1201.0	13239	-0.2	11958	0.0
	1129.0	12242	+0.2	11098	+0.1
	1063.0	11351.9	-0.1	10325.3	-0.1
Gold ingot	960.8	9990.9	-0.2	9137.9	0.0
Silver ingot	630.5	5919.5	0.0	5539.5	+0.1
	550.0	5009.7	0.0	4719.4	0.0
	500.0	4459.5	0.0	4219.8	0.0
	450.0	3920.7	0.0	3730.5	-0.1
	400.0	3397.1	0.0	3248.8	0.0
	350.0	2886.3	0.0	2776.6	0.0
	300.0	2391.9	0.0	2314.4	0.0
	250.0	1915.7	0.0	1865.7	0.0
	200.0	1462.7	0.0	1434.7	0.0
	180.0	1288.1	+0.1	1267.7	0.0
	160.0	1118.7	0.0	1104.0	0.0
	140.0	954.7	0.0	944.8	0.0
	120.0	795.4	+0.1	790.9	0.0
	100.0	644.2	0.0	643.2	-0.2
	80.0	498.4	+0.1	498.9	0.0
Resistance thermometer	60.0	360.4	+0.1	362.3	+0.1
	40.0	231.1	0.0	233.3	0.0
	20.0	110.7	+0.1	112.5	-0.1
	0.0	0.0	0.0	0.0	0.0

* The temperature differences are (observation - reference table).

It was mentioned earlier in this paper that the platinum, 10% rhodium-platinum thermocouple is the standard interpolation instrument for realizing the International Temperature Scale over the range from 630.5°C. to $1,063^{\circ}\text{C.}$ It is of interest to compare the results of interpolated values obtained by the specified method using the quadratic equation based on the E.M.F. values at the freezing

points of antimony, silver and gold with those given by the interpolation methods described in the Appendix. Taking the values of 5,539.5, 9,137.9 and 10,325.3 μ V. for the temperatures 630.5°, 960.8° and 1,063° C. as given in Table 1 for the 10% thermocouples, the following equation is derived:

$$E = -315.3 + 8.2305t + 0.001674t^2,$$

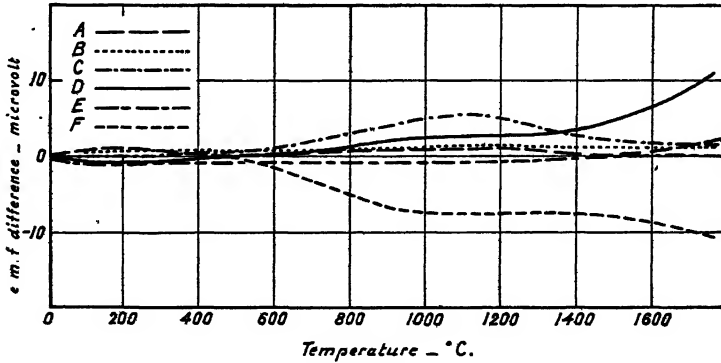


Figure 5. Departures of E.M.F. of each couple from mean of six 13% couples of make A.

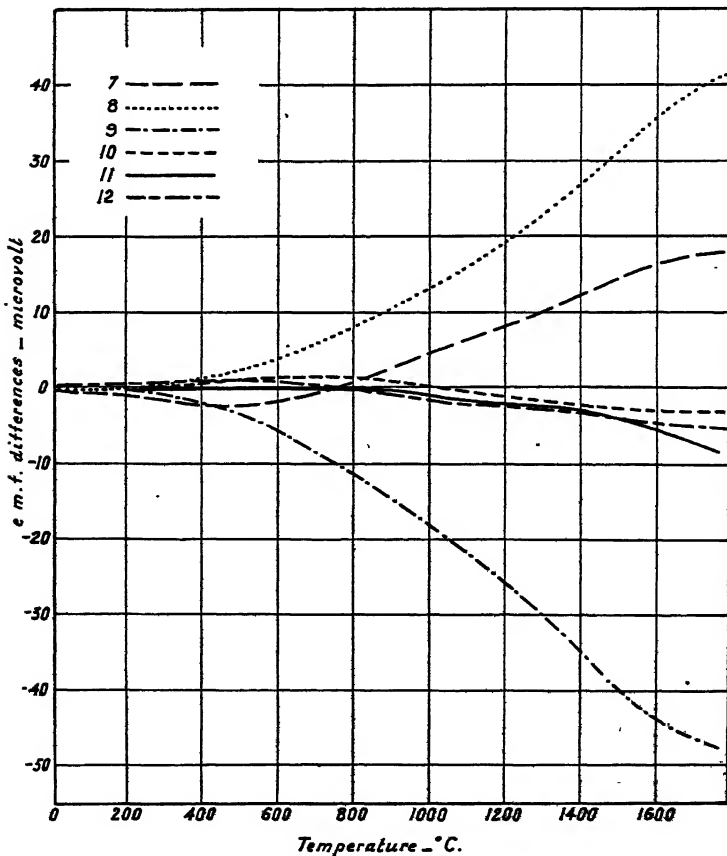


Figure 6. Departures of E.M.F. of each couple from mean of six 13% couples of make B.

Table 2
Reference Table for Platinum, 10% Rhodium-platinum Thermocouples, 1949

This Table is based on the mean calibration of twelve thermocouples, six from each of two manufacturers. The calibration was obtained by comparison with the resistance thermometer, by observations at the freezing points of silver, gold and platinum and by comparison with the optical pyrometer. The temperature scale used is the International Temperature Scale 1948 which, in the range covered by the Table, is based on the following fixed points:—

Melting point of ice 0° c.
Boiling point of water 100° c.
Boiling point of sulphur 444.60° c.
and from 1,063° c. upwards on the Planck law of radiation with a value of 1.438 cm. degrees for the constant C_2 .

Temperature in ° c. (Int. 1948). E.M.F. in microvolts (absolute). Cold junctions at 0° c.

t(°C.)	0	100	200	300	400	500	600	700	800	900	1000	1100	1200	1300	1400	1500	1600	1700	t(°C.)
0	0	642	1435	2314	3249	4220	5226	6265	7339	8446	9591	10757	11946	13155	14371	15580	16770	17922	0
10	55	716	1520	2406	3344	4319	5329	6371	7448	8559	9707	10875	12066	13277	14492	15700	16887	18035	10
20	112	791	1605	2498	3440	4419	5432	6477	7558	8672	9823	10993	12187	13399	14613	15820	17004	18147	20
30	172	867	1691	2590	3537	4519	5535	6584	7668	8786	9939	11111	12307	13520	14734	15940	17120	18259	30
40	233	945	1778	2683	3633	4619	5638	6691	7778	8900	10055	11229	12428	13642	14855	16060	17236	18371	40
50	297	1024	1866	2777	3720	4719	5742	6798	7888	9014	10172	11348	12549	13764	14976	16179	17351	18482	50
60	363	1104	1954	2871	3827	4820	5846	6906	7999	9129	10289	11467	12670	13886	15097	16298	17466	18592	60
70	430	1185	2043	2965	3925	4921	5950	7014	8110	9244	10406	11586	12791	14007	15218	16416	17581		70
80	499	1268	2133	3059	4023	5022	6055	7122	8222	9359	10523	11706	12912	14128	15339	16534	17695		80
90	570	1351	2223	3154	4121	5124	6160	7230	8334	9475	10640	11826	13034	14250	15469	16652	17809		90
100	642	1435	2314	3249	4220	5226	6265	7339	8446	9591	10757	11946	13155	14371	15580	16770	17922		100

Freezing point of silver
Freezing point of gold

0° c.
100° c.
444.60° c.

960.8° c.
1,063° c.

Table 3
Reference Table for Platinum, 13% Rhodium-platinum Thermocouples, 1949

This Table is based on the mean calibration of twelve thermocouples, six from each of two manufacturers. The calibration was obtained by comparison with the resistance thermometer, by observations at the freezing points of silver, gold and platinum and by comparison with the International Temperature Scale 1948 which, in the range covered by the Table, is based on the following fixed points:—

Melting point of ice 0° C. Freezing point of silver 960.8° C.

Boiling point of water 100° C. Freezing point of gold 1,063° C.

Boiling point of sulphur 444.60° C.

and from 1,063° C. upwards on the Planck law of radiation with a value of 1.438 cm. degrees for the constant C_1 .

Temperature in ° C. (Int. 1948). E.M.F. in microvolts (absolute). Cold junctions at 0° C.

t (°C.)	0	100	200	300	400	500	600	700	800	900	1000	1100	1200	1300	1400	1500	1600	1700	t (°C.)
0	0	644	1463	2392	3397	4460	5571	6735	7952	9209	10510	11850	13222	14617	16039	17463	18855	20202	0
54	54	75	89	97	104	109	114	120	124	128	133	136	139	141	143	141	137	132	10
57	57	77	89	98	3501	4569	5685	6855	8076	9337	10643	11986	13361	14758	16182	17604	18992	20334	10
111	111	79	91	99	3605	4679	5799	6975	8200	9465	10776	12122	13500	14899	16325	17745	19128	20466	20
170	170	81	91	100	3710	4789	5914	7095	8325	9594	10909	12258	13639	15041	16468	17885	19264	20598	30
231	231	81	93	100	3815	4899	6030	7216	8450	9723	11042	12395	13778	15183	16610	18025	19399	20729	40
295	295	83	94	101	3921	5010	6146	7338	8576	9853	11176	12532	13917	15325	16753	18165	19534	20860	50
361	361	84	94	102	4028	5121	6263	7460	8702	9983	11310	12669	14057	15468	16895	18304	19668	20990	60
429	429	86	95	102	4135	5233	6380	7582	8828	10114	11445	12807	14196	15610	17038	18442	19802		70
499	499	86	96	103	4243	5345	6498	7705	8955	10246	11580	12945	14336	15753	17180	18580	19936		80
571	571	88	97	103	4351	5458	6616	7828	9082	10378	11715	13084	14476	15896	17322	18718	20069		90
644	644			3397	4460	5571	6735	7952	9209	10510	11850	13222	14617	16039	17463	18855	20202		100

which gives values of E.M.F. of 6,799, 7,340 and 7,890 μV . at 750°, 800° and 850° c. respectively. These are in excess of the values in Table 1 by only 1, 1 and 2 μV . respectively.

Another point of interest is to observe the measure of agreement given by the tables with the International Temperature Scale criteria for the 10% type of couple to establish its suitability for use as a standard instrument. The criteria are

$$E_{\text{Au}} - E_{\text{Ag}} = 1,185 + 0.158(E_{\text{Au}} - 10,310) \pm 3 \mu\text{V}.$$

and

$$E_{\text{Au}} - E_{\text{Sb}} = 4,776 + 0.631(E_{\text{Au}} - 10,310) \pm 5 \mu\text{V}.,$$

where E_{Au} , E_{Ag} and E_{Sb} are the E.M.F.'s in microvolts at the freezing points of gold, silver and antimony. These criteria yield values on the basis of $E_{\text{Au}} = 10,325.3$ from Table 1 of 1,187.2 and 4,784.9, and the tables give 1,185.9 and 4,783.9, the agreement being well within the prescribed limits.

In the introduction to this paper it was mentioned that the reference table for 13% couples over the range from 1,400° to 1,750° c. was published in 1943, and at that time a full comparison of optical pyrometer and thermocouples could not be made. When this table has been corrected to the 1948 temperature scale and changed to absolute microvolts, the differences from the values in Table 3 shown in Figure 7 are obtained. The 1943 table was based on thermocouples having a mean E.M.F. at the palladium point of 18,198 absolute microvolts. The corresponding value for the present tables is 18,193, so that the tables refer to thermocouples of approximately the same E.M.F. The differences shown in Figure 7 nowhere exceed the equivalent of 1° c.

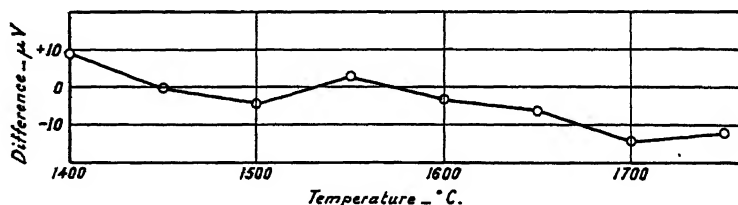


Figure 7. Difference between present reference table and 1943 table.
(Ref.: *J. Iron Steel Inst.*, No. 1, 1943, p. 205.)

ACKNOWLEDGMENTS

The work described above has been carried out as part of the research programme of the National Physical Laboratory, and this paper is published by permission of the Director of the Laboratory. The author desires to acknowledge the assistance rendered by Mr. L. H. Pemberton who carried out much of the observational work and by Mr. J. A. Hall for comments and suggestions.

REFERENCES

- BARBER, C. R., 1943, *J. Iron Steel Inst.*, No. 1, 205.
COMITÉ INTERNATIONAL DES POIDS ET MESURES, 1948, *Échelle Internationale de Température, Procès-Verbaux des Séances*, t. xxi.
GRACE, A., and HALL, J. A., 1943, *J. Sci. Instrum.*, 20, 60.

APPENDIX

Computation of the Reference Tables by L. Fox
Mathematics Division, National Physical Laboratory

There is no simple standard method for producing a smooth table, linearly interpolable, from a set of quantities obtained by physical measurement and given at unequal intervals of the argument. The lack of statistical knowledge of the conditions of the experiments, the weights to be attached to the various observations, and a probable mathematical relation between the functions concerned suggest that graduation in the actuarial and statistical sense is here impossible and even superfluous. The only criterion is that the final table should be smooth, and that the tabular entries corresponding to the arguments of the given data should differ as little as possible from the observed quantities.

The method adopted here is based on this criterion, and consisted of four stages:

(i) From standard processes of interpolation, using differences and divided differences where necessary, values of the E.M.F. were obtained at equal intervals of 50° C. over the whole temperature range. The differences will not be smooth, but after a certain order will oscillate about a mean value which is approximately zero. Differences of orders higher than this are neglected in the interpolation process. In the present case the first neglected difference was the fourth.

(ii) The values so obtained at equal intervals were differenced, and again the differences of some order showed oscillation about zero. This difference was smoothed, by a kind of relaxation process, making suitable changes in the functional values to reduce the differences to very small numbers oscillating about zero. Some skill is required here to produce the desired effect with the minimum functional change.

(iii) The smoothed values were then interpolated to fifths on the National Accounting Machine, the resulting tables having an argument interval of 10° C. and being linearly interpolable.

(iv) As a final check, the temperature arguments corresponding to the observed voltages were obtained by linear inverse interpolation and compared with the experimental figures. The discrepancies are shown in Table 1, and are seen to lie well within the possible limits of experimental error. This provides a complete verification of the adequacy of the process.

The work described above was carried out in the Mathematics Division of the National Physical Laboratory.

A Joule-Thomson Cascade Liquefier for Helium

By J. ASHMEAD

Royal Society Mond Laboratory, Cambridge

*Communicated by the Low-Temperature Group, read at Group Meeting on 7th December 1949;
MS. received 23rd January 1950.*

ABSTRACT. A helium liquefier has been constructed with an output of 3.8 litres per hour and a starting time of 70 minutes. The liquid hydrogen needed to cool the helium below its inversion temperature is provided continuously by a hydrogen liquefier within the same machine. The advantages of the machine are (i) short starting time, (ii) simplicity of operation and ease of servicing, (iii) closed hydrogen circuit.

§ 1. INTRODUCTION

BEFORE the subject of the present paper is discussed, it seems desirable to put on record the performance of its predecessor in the Mond Laboratory, the liquefier built by Kapitza (1934). Exact data for the years before the war are not readily available, but it is known that up to September 1939, the liquefier had been used on more than 350 days, and probably produced about 2,000 litres of liquid helium. Between December 1945 and May 1949, 216 runs were made and helium for 554 experiments was produced. This is a very fine record for a machine which was really intended as an experimental prototype rather than a final model. The starting time of the liquefier (using, of course, pre-cooling with liquid nitrogen) was only 50 minutes, and the rate of making liquid was 1.7 litres per hour.

By 1947, it became clear that the larger cryostats and more ambitious research programmes of the post-war years made a liquefier of about twice the existing capacity desirable. Since careful consideration was given to the possible alternatives, it may be useful to record the reasons which led to the solution adopted, since some other laboratories may be influenced by similar considerations. The new liquefier had to satisfy as many as possible of the following conditions:

- (i) Its output should be about 4 litres per hour.
- (ii) If possible, the existing compressors and gas-holders should be used.
- (iii) The apparatus should be as safe as possible.
- (iv) The starting time should be short.
- (v) Simplicity in running and ease in servicing are desirable even at the expense of a slight reduction in efficiency or an increase in the consumption of liquid nitrogen.

At first, the natural tendency was towards a re-designed liquefier of the expansion engine type. However, it appeared quite clear that no great increase in output was to be expected as long as the 'through-put' of the existing compressor was used. All that could be done was to produce a machine which might be a little more efficient than the Kapitza machine, and would be a good deal easier to service. To obtain the desired output, either a larger compressor would have to be installed, or else extra cooling would have to be provided by a subsidiary hydrogen circuit, as in the Moscow Kapitza machine.



In common with many other cryogenic laboratories the Mond Laboratory possesses a hydrogen liquefier, with a compressor and arrangements for storing the pure gas. It was therefore reasonable to consider at this stage those methods of helium liquefaction which demanded a supply of liquid hydrogen. The high thermodynamic efficiency of a liquefier of the Simon expansion type makes it very attractive in those cases where all the helium required for a day's experimenting can be produced in a single expansion, but if a number of successive compressions and expansions have to be made to provide all the liquid wanted, it seems less certain whether the *average* rate of production of liquid will be higher than with a continuously running, though less efficient, machine. Moreover, the Mond Laboratory does not possess that experience of small Simon liquefiers which other laboratories enjoy, so that the design of a large machine of that type would have been rather more of a new departure than most of the other possible devices.

The other important existing type of helium liquefier is that employing straightforward Joule-Thomson expansion of helium pre-cooled by liquid hydrogen, such as that in the Kamerlingh Onnes Laboratory at Leiden. Examination of the performance of this machine showed that it should be easy to obtain the output required with the helium compressor available. On the other hand, such a scheme had two important disadvantages. It involved the use of liquid hydrogen, with its attendant dangers, and it entailed two starting times instead of one. The second of these disadvantages could be overcome by making the hydrogen on a previous day and storing it (see for example Jones, Larsen and Simon 1948), so that the liquid helium could be produced fairly easily on the days when it was needed, but it seemed as though there might be another and more attractive possibility.

The suggestion was that a hydrogen and helium liquefier should be combined within the same container. The liquid hydrogen produced would never be decanted from the machine, but would serve only to cool the helium below its inversion temperature. A small liquefier of this type was described by Seiler (1941) but it differs in so many respects from the present machine that the two are hardly comparable. The advantages of a liquefier of this type are appreciable. The hydrogen circuit is completely closed and the gas therefore needs purification only very infrequently. Further, since no liquid hydrogen ever leaves the apparatus, the explosion hazard is considerably reduced compared with a system in which liquid or gas is frequently handled. Again, the combination of the two liquefiers in one envelope enables a very good overall starting time to be obtained—certainly less than the sum of the two separate starting times. Experience with the existing hydrogen liquefier led us to believe that a liquefier of the Joule-Thomson type would be straightforward to run. These arguments resulted in the choice of the type of liquefier described in this paper, and subsequent experience of the completed machine has, in our opinion, justified the decision made.

§ 2. GENERAL DESIGN OF THE LIQUEFIER

Before the details of the various components of the machine are examined, it is convenient to describe the general arrangement of the liquefier illustrated in the Plate. The distance from the plate near the top of the liquefier to the floor is about five feet. High pressure hydrogen (150 atmospheres) and high

pressure helium (25 atmospheres) are cooled in the top exchanger by returning hydrogen, returning helium, and nitrogen from the liquid nitrogen bottle just below the exchanger. The liquid nitrogen in this bottle boils at about 67°K . under reduced pressure and the high pressure hydrogen and helium pass through tubes soldered to the outside of the bottle and leave it at a temperature within one or two degrees of its own. In the next exchanger the high pressure gases are cooled by the returning gases, and after this exchanger the hydrogen expands at a Joule-Thomson valve to a pressure of 1.5 atmospheres. Part of the hydrogen is liquefied and collects in the upper hydrogen bottle and the remainder of the gas passes back up the heat interchangers. A small fraction (about one-fifth) of the liquid passes through a transfer valve into the lower hydrogen bottle where it boils at 6 cm. pressure, producing a temperature of 14°K . The gas from this lower bottle passes only through the short exchanger situated between the two hydrogen bottles, and then goes straight to room temperature via a wide-bore tube. This causes a slight loss of thermal efficiency but solves the problem of producing a high pumping speed at the lower bottle. Two hydrogen bottles are used rather than one because if all the gas were expanded to 6 cm. pressure, the speed of the vacuum pump would have to be about 25 times as great. The high pressure helium is cooled by passage through the tubes soldered to the outside of both hydrogen bottles, and by the hydrogen gas in the intermediate exchanger.

The helium thus enters the final exchanger at 14°K ., and after passing through the exchanger it expands at a Joule-Thomson valve where part of it liquefies and collects in the large copper bottle at the bottom of the machine. From there, the liquid is siphoned into the flasks through the vacuum jacketed tube which can be seen near the top right-hand corner of the photograph.

An important feature of the design of this machine lies in the precautions taken to reduce the starting time. During the early stages of operation the hydrogen and helium do not return up the exchangers but pass straight from the bottles back to the gas-holders via 'circulation valves' on the top of the liquefier. This causes a certain waste of cold in the top exchanger, and hence a greater consumption of liquid nitrogen, but it results in a more rapid cooling of those parts of the machine below the liquid nitrogen bottle. The wide-bore tubes to the circulation valves from the hydrogen and helium bottles are also connected via safety valves to the gas-holders in case a dangerous pressure builds up in either bottle.

Some of the more detailed features of the design may now be examined.

§ 3. THE HEAT INTERCHANGERS

The top three heat exchangers call for little comment. They consist essentially of bundles of tubes soldered together. The data for their design are taken from Jacobs and Collins (1940), and the actual method of construction is very similar to that employed by Blanchard and Bittner (1942). The bottom exchanger, on the other hand, is rather novel in design. At very low temperatures the high density of the helium allows it to flow turbulently even through very small bore tubes. An ideal heat exchanger would therefore consist, on the low pressure side, of a bundle of some fifty tubes of 1mm. bore, all in parallel. This effect can be produced by soldering to the high pressure tube a piece of crinkled copper foil, which is then wrapped round on itself and pushed inside a thin-walled

German silver tube. Thus the spaces between the crinkles form fine bore tubes giving very efficient heat transfer between the gas and the copper. The copper is near the maximum of its thermal conductivity in this temperature range, and so the cold is efficiently conducted to the central tube. To prevent longitudinal conduction, the copper foil is cut into pieces 6 cm. long. The form eventually taken by this heat interchanger is rather like a multiple trombone, and the first few sections of it can be seen surrounding the helium bottle at the left of the photograph.

§ 4. THE VALVES

The Joule-Thomson valves and the transfer valve are almost identical in design. The threaded part of the valve is at room temperature, near the top plate rather than close to the expansion orifice, so as to reduce the possibility of a valve seizing up because of impurities. The valves are so constructed that the whole valve, including the seating, can be withdrawn through the top plate after only two simple soldered joints have been broken. The design adopted for the circulation valves is a development from the tap used in the Kapitza liquefier for drawing off the liquid helium (Kapitza 1934).

§ 5. THE SIPHON TUBE

The liquid helium is drawn off from the helium bottle through a German silver tube over two metres long, 3 mm. in diameter and 0.1 mm. in wall thickness. This tube is, of course, protected by a vacuum jacket and a copper radiation shield, which is thermally bonded to the helium return tubes near the liquid nitrogen bottle.

§ 6. RADIATION SHIELD

The part of the liquefier below the nitrogen bottle is protected from the influx of room temperature radiation by a copper shield about 4 mm. thick, which is cooled by conduction from a subsidiary nitrogen bottle at its lower end.

§ 7. CASING

The liquefier is surrounded by a brass case with a Gaco ring seal to the brass plate from which the liquefier hangs. A vacuum of better than 10^{-5} mm. of mercury is maintained inside the casing by a diffusion pump of speed 20 litres per second. Even when the casing is shut off from the pump by a high vacuum valve, the thermal isolation is very good. After the liquefier is closed down, although no more liquid nitrogen is siphoned into the shield bottle, the rate of evaporation of helium is only 0.1 litre per hour and there is always some liquid hydrogen left in the hydrogen bottle the next day.

§ 8. INDICATORS

Besides the usual complement of pressure gauges, the liquefier is equipped with simple gas thermometers at the hydrogen and helium expansion valves. The levels of the liquids in the various bottles are observed by measuring on an inclined butyl phthalate manometer the pressure difference between the gas above the free surface in the bottle, and that in a tube connected to the bottom of the bottle. This second tube is curled in a horizontal plane just below the bottle, so that the free surface in the tube is always in the horizontal part of the tube,

even if the heat flow down the tube varies appreciably. The butyl phthalate gauges are fitted with special self-returning safety traps, to guard against sudden variations of pressure, and they are by-passed by suitable valves until conditions are steady. Despite the very small pressure differences available owing to the low density of liquid hydrogen, the level indicators work very satisfactorily.

§ 9. GENERAL

A great deal of attention has been paid in the design of this machine to ease of servicing, and no attempt has been made to make the design compact by enclosing one part of the liquefier inside another. This principle has produced a rather tall machine, but one in which every vital part can be easily exposed for examination and repair. A pit has been dug in the floor under the liquefier so that the outer casing and the radiation shield can be removed with the minimum of trouble. About 20 minutes is required to remove the case and half an hour to disconnect the radiation shield, and very little longer to replace them. All the components hang from the top plate of the machine, which is supported by a cantilever so that everything is accessible.

§ 10. PERFORMANCE

Liquid hydrogen is produced about 30 minutes after the machine has started, and in a further 40 minutes the liquefier has reached equilibrium and is producing liquid helium. During this starting period 15 kg. of nitrogen is used in the main bottle and 10 kg. to cool down the shield. Once the shield is cold, very little more nitrogen is needed to keep it cold. In the equilibrium state, the through-put of hydrogen is 21 m³/hour at a pressure of 150 atmospheres, and the helium through-put is 17 m³/hour at 25 atmospheres. The consumption of liquid nitrogen is 8 kg/hour and the output of liquid helium is 3.8 litre/hour. If the machine is operated as a hydrogen liquefier, it produces 5 litre/hour, which is about 80% of the theoretical yield. The helium part of the circuit usually gives about 75% of the theoretical optimum, though as much as 85% has sometimes been obtained. The lower efficiency is accompanied by an increased pressure drop, and is probably due to a slight 'plating' of the exchangers by residual impurities.

ACKNOWLEDGMENTS

During the earlier stages of the design, I had many helpful discussions with Professor J. F. Allen, and Dr. D. Shoenberg encouraged me during the later phases of the work. A particular acknowledgment is due to Mr. F. Sadler who freely placed his experience of earlier machines at my disposal and was responsible for much of the detailed design work. The greater part of the actual construction was carried out most efficiently by Mr. E. Boud.

REFERENCES

- BLANCHARD, E. R., and BITTNER, H. W., 1942, *Rev. Sci. Instrum.*, **13**, 394.
 JACOBS, R. B., and COLLINS, S. C., 1940, *J. Appl. Phys.*, **11**, 491.
 JONES, G. O., LARSEN, A. H., and SIMON, F. E., 1948, *Research*, **1**, 420.
 KAPITZA, P., 1934, *Proc. Roy. Soc. A*, **147**, 189.
 SEILER, K., 1941, *Ann. Phys., Lpz.*, **39**, 129.

Time Decrease of Magnetic Permeability in Alnico

BY R. STREET AND J. C. WOOLLEY

The University, Nottingham

Communicated by L. F. Bates; MS. received 23rd November 1949

ABSTRACT. The formal theory, based on the concept of thermal activation of domain processes, previously proposed to account for the phenomenon of magnetic viscosity, has been extended to describe time decrease of permeability. Within certain limits the predictions of the theoretical analysis are verified by measurements made on time decrease of permeability using an alnico rod specimen. It is shown that the quantitative results of these measurements are in good agreement with those obtained from the magnetic viscosity experiments. The possible physical mechanisms responsible for both magnetic viscosity and time decrease of permeability in alnico are discussed.

§ 1. INTRODUCTION

MEASUREMENTS of incremental permeability μ_i of a ferromagnetic material at any point on the magnetization curve are made by finding the change in magnetic induction ΔB produced by a known change in the magnetic field ΔH . For the majority of materials the value of μ_i so obtained is independent of the time which elapses (*resting time*) between the establishment of H and the application of the increment ΔH . However, when the incremental permeability of certain magnetic materials of low coercivity is examined in this way, μ_i is found to decrease as the resting time increases; this phenomenon is known as time decrease of permeability (Webb and Ford 1934). Snoek (1941) has shown that the effect may be observed in soft iron specimens only if they contain small quantities of carbon ($\sim 0.008\%$) or nitrogen ($\sim 0.007\%$) as impurity. He considers that the observed time decrease of permeability is due to a diffusion process of the impurity atoms within the iron lattice. Snoek has also investigated the magnetic viscosity exhibited by such soft iron specimens and has shown that there is an intimate relationship between the phenomena of magnetic viscosity and time decrease of permeability.

In this communication a report is given of the observations which have been made on time decrease of permeability in alnico, a high coercivity permanent magnet alloy. The related phenomenon of magnetic viscosity in alnico has already been investigated (Street and Woolley 1949 b, referred to below as I), and a formal theory of the phenomenon given. This theory was based on the concept of the activation of domain processes by thermal energy. In the first part of the present paper, this formal theory is developed to cover the case of time decrease of permeability, and discussion of the underlying physical mechanism is deferred until a later section.

§ 2. THEORY

As in I the underlying assumption is made that at some stage in the magnetization of a ferromagnetic substance a certain number of domains have their magnetization vectors in positions of metastable equilibrium. Hence, the domain vector orientation is scarcely altered by a small increase in the domain

energy, but a greater energy increase initiates a transition to stable equilibrium which, in turn, involves a discontinuous, irreversible change in vector orientation. Consequently, the process envisaged is essentially one of activation, and the critical value of the energy required to produce discontinuous rotation will be referred to below as the activation energy of the domain. This energy can be supplied by an increase in the applied magnetic field in which case an increase in the magnetization of the specimen is observed. The experimental results given in I, which show that there is a spontaneous increase in the magnetization of the specimen in a steady applied field, are consistent with the assumption that the required activation energy may be supplied by random thermal fluctuations within the material. This leads to the prediction of time decrease of permeability. For, suppose that the applied field H (defined in §1) is of such magnitude that spontaneous increase of magnetization occurs, then during the resting time t a certain number of domains are activated by thermal energy. Therefore, the number of domains which are activated by the eventual application of ΔH decreases as a function of t and, consequently, the corresponding value of ΔB decreases in a similar way, leading to an observed time decrease of permeability.

The quantitative analysis of this process is as follows. A steady field H is applied to the specimen at zero time. Let N be the number of domains having activation energies lying between E and $E + dE$ at time t .

Then,

$$\frac{dN}{dt} = -CN \exp\left(-\frac{E}{kT}\right) \quad \dots\dots(1)$$

where C is a constant, k is Boltzmann's constant and T is the absolute temperature of the specimen. In the derivation of this equation it is assumed that (a) successful activations do not produce additional metastable states with activation energies in the range from E to $E + dE$ and (b) each domain may be characterized by a constant activation energy independent of the state of the neighbouring domains, i.e. the magnetic interaction between domains is assumed to be negligible. Integrating equation (1) we have

$$\ln N = -\lambda t + \text{const.} \quad \dots\dots(2)$$

where $\lambda = C \exp(-E/kT)$.

In a general analysis, N should be taken as any arbitrary function of the activation energy, E . However, some definite form of this function is required for the further development of the analysis. In I the function was given the simple form defined as follows: (a) the domains have activation energies lying only in the range from zero to an upper limit, E_0 , and (b) within this range, N at time $t = 0$ is independent of E and has a value $p dE$, where p is the distribution constant. As far as the phenomenon of magnetic viscosity is concerned, the result of the theoretical analysis using this simplification is in very good agreement with our experimental observations. This form of the distribution is therefore retained in the following analysis of time decrease of permeability. Thus equation (2) becomes

$$N = p \exp(-\lambda t) dE. \quad \dots\dots(3)$$

When ΔH is applied, the magnetization vectors of all domains having activation energies below some value E' will rotate discontinuously, the actual value of E' being directly determined by ΔH . The resultant total change in the intensity

of magnetization of the specimen will consist of two parts, (a) the contribution of the discontinuous irreversible movements, which will be denoted by $\Delta I'$, and (b) the contribution of the reversible movements of the magnetization vectors of all the domains of the material. In order to determine the contribution $\Delta I'$ it is necessary to find the total number of domains, N_{tot} , having activation energies lying between zero and E' at time t after the field H has been established. This is obtained from equation (3):

$$N_{\text{tot}} = \int_0^{E'} p \exp(-\lambda t) dE.$$

Thus, if \bar{i} is the average value of the change in the resolved part of the domain magnetization along the direction of the applied field due to the activations, then

$$\Delta I' = \bar{i} p \int_0^{E'} \exp(-\lambda t) dE. \quad \dots\dots(4)$$

Changing the variable to λt , equation (4) becomes

$$\begin{aligned} \Delta I' &= -\bar{i} p k T \int_{Ct}^{\lambda' t} \frac{1}{\lambda' t} \exp(-\lambda t) d(\lambda t) \\ &= \bar{i} p k T [\text{Ei}(-Ct) - \text{Ei}(-\lambda' t)], \end{aligned} \quad \dots\dots(5)$$

where $\lambda' = C \exp(-E'/kT)$, and $-\text{Ei}(-x)$ is the exponential integral, defined (Glaisher 1870) by

$$\int_x^\infty \frac{1}{u} \exp(-u) du.$$

The total change in the intensity of magnetization due to the application of ΔH will contain a contribution from reversible rotations of the domains, and this may be accounted for by the addition of a term $f(\Delta H)$ to the right-hand side of (5) which now becomes

$$\Delta I_{\text{tot}} = \bar{i} p k T [\text{Ei}(-Ct) - \text{Ei}(-\lambda' t)] + f(\Delta H). \quad \dots\dots(6)$$

In the derivation of equation (6) it has not been necessary to make any assumptions concerning the ranges of t and E' . A general equation of very similar form to that of equation (6) has been derived in the case of magnetic viscosity (Street and Woolley 1949 c), viz.

$$\Delta I = \bar{i} p k T [\text{Ei}(-\lambda_0 t) - \text{Ei}(-Ct) + E_0/kT], \quad \dots\dots(7)$$

where ΔI is the spontaneous increase in the intensity of magnetization after the resting time t , and $\lambda_0 = C \exp(-E_0/kT)$. It has been found possible to make a direct check of equation (7) in its most general form since it is in satisfactory agreement with the experimental observations made by Snoek (1947) on magnetic viscosity in Mn-Zn ferrite.

A general plot of equation (7) shows that over a range of resting times t defined by the inequalities $t_0 > t > t_c$, where

$$\left. \begin{aligned} |\text{Ei}(-Ct_c)| &\ll |\text{Ei}(-\lambda_0 t_0)| \\ \text{and } \lambda_0 t_0 &\ll 1 \end{aligned} \right\} \quad \dots\dots(8)$$

the value of ΔI is given by

$$\Delta I = \bar{i} p k T \ln t + \text{const.}$$

The convenient range of times during which experimental observations on magnetic viscosity in alnico were made was found to lie well within the limits of t defined immediately above. The experimental results were thus satisfactorily represented by the preceding $\ln t$ relation.

Subject to the limitations already imposed in the derivation of equation (6), the latter should give a satisfactory description of time decrease of permeability. Since the experimental results on this phenomenon given in the present paper were obtained for the same alnico specimen over the same time ranges which were employed in the magnetic viscosity experiments, (6) is now considered over a similarly limited range of times, viz. $t_0' > t > t_c'$ where

$$\left. \begin{aligned} & |Ei(-Ct_c')| \ll |Ei(-\lambda't_0')| \\ & \text{and } \lambda't_0' \ll 1. \end{aligned} \right\} \dots\dots(9)$$

Within the range of times so specified equation (6) becomes

$$\Delta I_{\text{tot}} = -\bar{i}pkT(\ln \lambda't + \text{const.}) + f(\Delta H). \dots\dots(10)$$

The limiting values of t_0' and t_c' defined by (9) now depend on the value of E' , and hence on the corresponding value of the applied magnetic field increment ΔH . From the experiments on magnetic viscosity in alnico it would be expected that the time range from t_c' to t_0' would include the times conveniently available for experiment (from 30 to 3,000 sec.) when E' is comparable with E_0 , i.e. when relatively large values of ΔH are used in the experiments.

In experimental work designed to test the predictions of equation (10) certain variables may be conveniently held constant. For example, if the increment in applied magnetic field, ΔH , is maintained at a constant value in a series of experiments, since E' is now constant equation (10) may be written

$$\Delta I_{\text{tot}} = -\bar{i}pkT(\ln t + \text{const.}) + f(\Delta H). \dots\dots(11)$$

Again, when the time t and the absolute temperature T are held constant, the variation of ΔI_{tot} as a function of E' is

$$\Delta I_{\text{tot}} = \bar{i}pE' + \text{const.} + f(\Delta H). \dots\dots(12)$$

The present experimental work was directed towards the verification of equations (6), (11) and (12) and the evaluation of the constants characteristic of the domain processes involved in the magnetization cycle.

§ 3. APPARATUS AND METHOD

The rod of alnico employed in the present investigations was that used in I where details of its composition, dimensions and heat treatment are given. The specimen, on which a search coil was wound, was supported symmetrically either in an electrically heated tubular oven, the heater of which was wound non-inductively and fed with D.C. or, when temperatures below that of the room were required, in a Dewar vessel. The required external magnetic field was applied by means of a surrounding water-cooled solenoid capable of producing a maximum field of 1,100 oersteds. The current flowing in the solenoid, supplied by a 280 volt battery of accumulators, was measured by a potentiometer method. The search coil was connected to a ballistic galvanometer with a standard mutual inductance included in circuit for calibration purposes. Two galvanometers were used, one of period 6 seconds and the other of period 16 seconds.

The temperature of the specimen was determined by a calibrated thermocouple attached to its mid-point, and measurements were made at fixed temperatures within the range -80°C. to 130°C. , the maximum temperature fluctuations about the steady values being approximately 2°C.

The general principle underlying the measurements has already been outlined in § 1. In the majority of the experiments, the steady applied field H was equal to the coercive field H_C of the specimen. Initially the alnico rod was brought into a cyclic state of magnetization and the apparatus was then allowed to reach thermal equilibrium with external demagnetizing field H_C applied to the specimen. Immediately before a measurement was made one more complete hysteresis cycle was traversed. After the waiting period t seconds had elapsed, the field increment ΔH was applied, and the resulting change in magnetic induction was measured by means of the ballistic galvanometer. It was found most convenient to apply the field increment ΔH to the specimen by passing a current through an auxiliary solenoid wound over the main solenoid. In this way it was possible to obtain values of ΔH which were consistent and reproducible to within $\frac{1}{4}\%$, a much better reproducibility than was obtained by a sudden increase of the current in the main magnetizing solenoid.

§ 4. RESULTS AND INTERPRETATION

Measurements were first made in an attempt to verify equation (6). A typical set of results of variation of ΔI_m with t is plotted logarithmically in Figure 1.

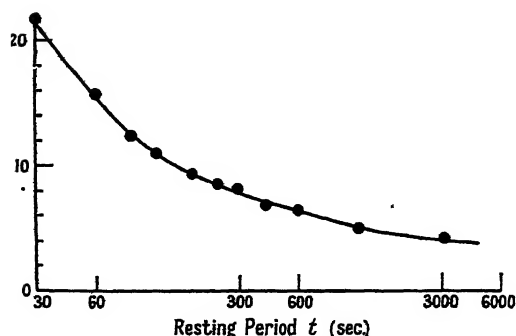


Figure 1. The variation of ΔI_m as a function of the resting period t . For each point the increment of the external field, $\Delta H = 6.7$ oersteds. Galvanometer period = 6 sec.

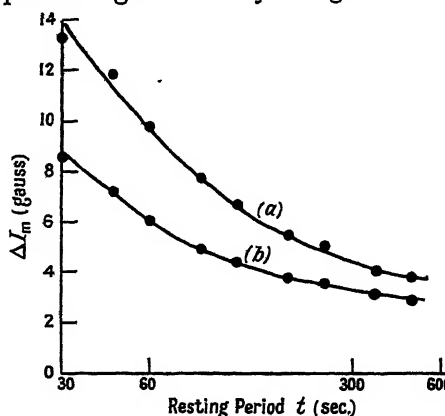


Figure 2. The variation of ΔI_m as a function of resting time, t sec., with $\Delta H = 4.3$ oersteds.

Curve (a) galvanometer of period 16 sec.
Curve (b) galvanometer of period 6 sec.

ΔI_m is used to denote the measured value of the change of magnetization of the specimen. The measurements were made at room temperature and the increment ΔH was 6.7 oersteds for every individual determination. It was found that the curve drawn through these experimental points could not be represented by a difference of exponential integral functions as predicted by equation (6). This apparent failure of equation (6) to describe the experimental results is explained by the fact that, in its derivation, only those contributions to the measured change of magnetization directly due to the application of ΔH were considered. In practice, however, magnetic viscosity will occur at a field strength of $H + \Delta H$,

i.e. immediately after the increment ΔH has been applied, and this will add a further contribution to the measured value of the change of induction because of the finite period of the ballistic galvanometer. Conclusive evidence of the existence of a magnetic viscosity contribution was obtained experimentally by making measurements on time decrease of permeability using two galvanometers with appreciably different periodic times, viz. 6 and 16 seconds. The results of measurements made at room temperature with $\Delta H = 4.3$ oersteds are shown in Figure 2. These two sets of measurements were made under identical conditions except for the two different galvanometers used, and it is seen that the values of the changes in the intensity of magnetization obtained using the 16-second galvanometer (curve (a)) are greater than the corresponding values obtained using the 6-second period instrument. Thus the longer the periodic time of the galvanometer the greater the contribution of the magnetic viscosity to the measured change in intensity of magnetization.

In order to take account of this magnetic viscosity correction it is necessary to introduce another term into the right-hand side of equation (6). A consideration of equation (3), which gives the number of domains having activation energies

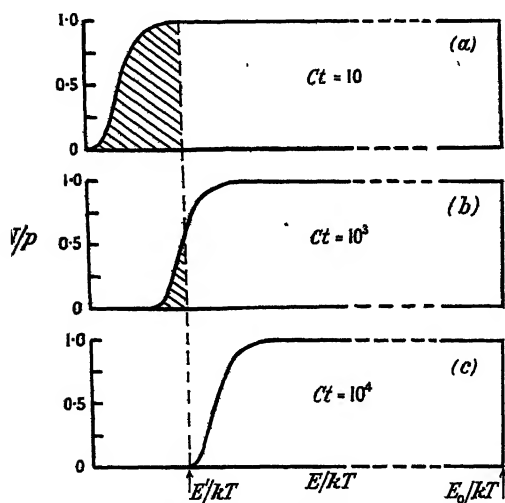


Figure 3. Variation of distribution function after various resting periods t . For explanation of shading and symbols see text.

lying between E and $E + dE$ as a function of E , shows the difficulty of estimating this correction term. The form of the required distribution functions may be calculated from equation (3), and the results of these calculations are shown diagrammatically in Figure 3, where N/p is plotted as a function of E/kT for the three values of Ct given for the individual curves. The domains represented by the shaded areas contribute to $\Delta I'$, whilst those represented by the unshaded areas are responsible for the disturbing magnetic viscosity effect. It is possible to obtain the analytical form of the distribution represented by the unshaded areas, but the resulting mathematical analysis of magnetic viscosity effects, assuming this form of distribution, is intractable, except for the case (a) of Figure 3, where the resultant distribution function at field $H + \Delta H$ has the simple form $N(E) = p = \text{constant}$. This corresponds to the case treated in I.

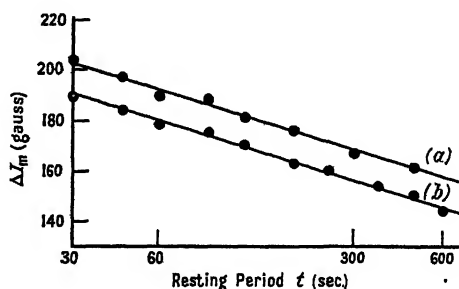


Figure 4. The variation of ΔI_m as a function of the resting time, t sec., with $\Delta H = 21.4$ oersteds.

Curve (a) using galvanometer of period 16 sec.
Curve (b) using galvanometer of period 6 sec.

Because of the limitations described above, an attempt was made to eliminate the contribution due to magnetic viscosity from the measured value of the change in induction by the following experimental method. A cascade of electro-mechanical relays was employed to disconnect the ballistic galvanometer from the search coil at various short times, δt (of the order of tens of milliseconds), after the increment ΔH had been applied. The observed change of induction for constant ΔH and constant resting period t was found to be a continuous function of δt . There was thus no justification for taking any particular value of the change of induction as corresponding to ΔI_{tot} . It was supposed that this result was due to (a) the time constant of the solenoid and battery circuit, and (b) the effects of eddy-current lag in the specimen, resulting in a relatively slow building up of the increment of field ΔH . Consequently, any calculation of the magnetic viscosity correction term which assumes an instantaneous application of ΔH is of little practical value.

Since the observed values of the change of induction contain an undetermined contribution due to magnetic viscosity, quantitative agreement between the experimental results and equation (6) cannot be expected. This difficulty is avoided to a considerable extent when, in any set of measurements, the contribution of magnetic viscosity is constant for all the individual measurements. It is shown below that this condition is satisfied if, after the application of the increment ΔH at time t , the *remaining* domains are characterized by a constant activation energy distribution p . For, from equation (1), the time rate of change of the intensity of magnetization at $t=0$ due to magnetic viscosity is given by

$$\begin{aligned} \left(\frac{dI}{dt}\right)_{t=0} &= \bar{i}pC \int_0^{E_0} \exp(-E/kT) dE \\ &= \bar{i}pkTC \left[1 - \exp\left(-\frac{E_0}{kT}\right)\right]. \quad \dots\dots(13) \end{aligned}$$

In I it was shown that for this alnico specimen $E_0 \gg kT$, a statement implicit in the first inequality of equation (8) above. Therefore, to a good approximation, equation (13) may be written as

$$\left(\frac{dI}{dt}\right)_{t=0} = \bar{i}pkTC.$$

Thus the contribution to the measured change of magnetization due to magnetic viscosity in a time Δt , determined by the time of swing of the galvanometer, is approximately $\bar{i}pkTC \Delta t$, which is a constant contribution for constant p . Reference to Figure 3 shows that the distribution function characteristic of the unactivated domains remaining after the application of ΔH will only be constant if the final conditions are as suggested by the curve (a). This means that for the maximum resting period to be used in the experiments the value of E' , and consequently ΔH , must be of a sufficiently large magnitude. Thus for these cases equation (6) may be written

$$\Delta I_m = \bar{i}pkT[\text{Ei}(-Ct) - \text{Ei}(-\lambda't)] + \text{const.} + f(\Delta H), \quad \dots\dots(14)$$

where ΔI_m denotes the value of the change of magnetization which should be obtained in experimental measurements. The effect of magnetic viscosity is now included in the constant term on the right-hand side of equation (14), and now this

constant depends not only on the domain constants but also on the constants of the ballistic galvanometer and the associated circuits. The limitations of this analysis now become apparent, for, owing to the inclusion of the indeterminate constant term in equation (14), only the slopes of the experimental curves may be used to determine the fundamental constants of the domain processes.

If the magnitude of ΔH is sufficiently large so that equation (14) holds, then, although the measured values of the change in magnetization should be different if observations are made with two ballistic galvanometers of different periodic times, the difference should be constant and independent of the resting period t . Experimental evidence for this conclusion is presented in the two sets of curves shown in Figures 2 and 4. In the former case, where the size of the field increment ΔH is 4.3 oersteds, the difference between the measured values with ballistic galvanometers of periods 6 seconds and 16 seconds is a function of the resting time t . However, if the experiment is repeated for a larger value of ΔH , namely 21.4 oersteds, the two sets of results differ only by a constant value. It will also be seen from Figure 4 that, with $\Delta H = 21.4$ oersteds, the inequalities (9) are satisfied, for a linear variation of $(\Delta I_m, \ln t)$ is obtained. The equation of this straight line is given by (11), after modification to take into account the contribution due to magnetic viscosity, i.e.

$$\Delta I_m = -\bar{i}pkT(\ln t - \text{const.}) + f(\Delta H) + \text{const.} \quad \dots\dots (15)$$

From equation (15) the slope of the $(\Delta I_m, \ln t)$ curve is $-\bar{i}pkT$. The magnitude of the slope of the lines shown in Figure 4 is 14.7 gauss, measured at a temperature of 14°C., which differs by less than 1% from the value of $\bar{i}pkT$ derived from the previous measurements on magnetic viscosity described in I.

The values of ΔI_m were determined as a function of time, using various fixed values of ΔH . These results are plotted in Figure 5 and the values of ΔH

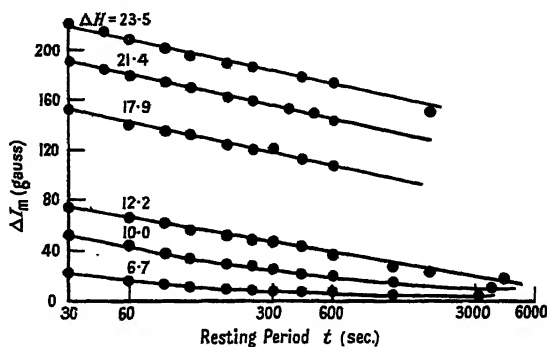


Figure 5. Variation of ΔI_m as a function of the resting period t for the various values of ΔH (oersteds) indicated on the curves. Galvanometer period = 6 sec.

used are indicated on the individual curves. It is apparent from this figure that when $\Delta H \geq 18$ oersteds then (a) ΔI_m is proportional to $\ln t$ for values of t from 30 to 600 seconds, and (b) the slopes of the individual lines are equal and have a common value 14.7 gauss; (temperature 14°C.). This is in agreement with the prediction of equation (15) that the slope of the $(\Delta I_m, \ln t)$ curve, i.e. $\bar{i}pkT$ should be independent of ΔH , for sufficiently large values of ΔH . Thus the

experimentally determined slope is again in good agreement with the results of the magnetic viscosity measurements. The family of curves in Figure 4 serves to define the ranges of ΔH and t for which the inequalities given in equation (9) are justified.

A further prediction of equation (15) is that, provided $\bar{i}p$ is independent of temperature, the slope of the $(\Delta I_m, \ln t)$ curve should be proportional to the absolute temperature of the specimen. Experiments were made to check this prediction by making measurements of ΔI_m as a function of resting time t when the specimen was maintained at temperatures of -78° , 14° , and 130° c. The same value of ΔH , 21.4 oersteds, was used in the experiments at the three different temperatures, and the results are plotted in Figure 6. The slopes of the experimental curves have the values 9.9, 14.7 and 20.6 gauss. Thus the ratio of slope

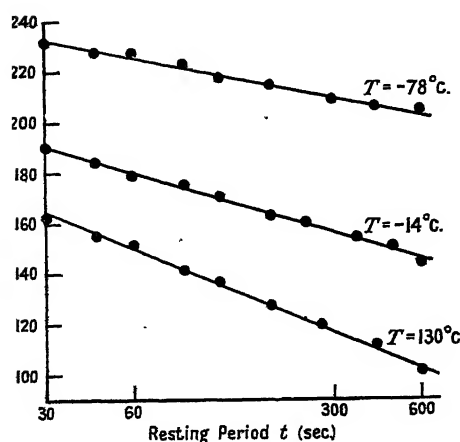


Figure 6. Variation of ΔI_m as a function of t with the specimen maintained at the different temperatures indicated on the curves, $\Delta H = 21.4$ oersteds. Galvanometer period = 6 sec.

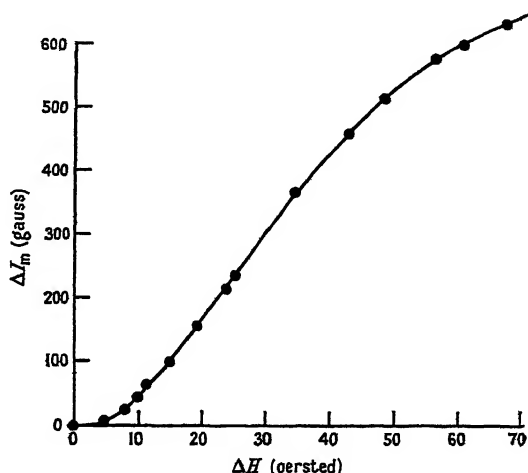


Figure 7. Variation of ΔI_m as a function of the change in the applied magnetic field ΔH for a constant resting period $t = 1$ minute. Galvanometer period = 6 sec.

to absolute temperature, i.e. $\bar{i}pk$, is constant, being equal to 0.051 gauss per degree Centigrade in each case, to within $\pm 1\%$, giving a value of $\bar{i}p$ equal to 3.7×10^{14} gauss per erg, which is in good agreement with the value of 3.6×10^{14} gauss per erg derived from the measurements on magnetic viscosity given in I.

The presence of the magnetic viscosity term precludes the possibility of determining the value of C from an analysis, based on equation (14), of the experimental results.

The experimental verification of equation (12) is not directly possible since the forms of $f(\Delta H)$, introduced to account for reversible changes, and $F(\Delta H) = E'$ are not known *ab initio*. However, measurements were made of ΔI_m as a function of ΔH with a resting period $t = 60$ seconds for each measurement; the temperature of the specimen was constant at 20° c. The results are plotted in Figure 7 and it will be seen that over a range of ΔH from 10 to 40 oersteds, ΔI_m is practically a linear function of ΔH . Qualitatively, with the constant distribution factor p assumed in the analysis, as indicated in Figure 3, it would be expected that $\Delta I'$, the contribution to the change in intensity of magnetization due to irreversible

activation of the domains, would reach a constant value when $E' = E_0$. This corresponds physically to the situation when all the irreversible domain changes are activated by the application of ΔH . The consequence is that a decrease in the slope of the $(\Delta I_m, \Delta H)$ curve would be expected for values of ΔH above that corresponding to E_0 . This is a possible explanation of the decrease of slope of the $(\Delta I_m, \Delta H)$ curve, for values of ΔH greater than 40 oersteds.

It is not possible to go further in the interpretation of the results shown in Figure 7 in terms of equation (12), without making further assumptions. The underlying physical mechanism responsible for the phenomena will now be considered.

§ 5. DISCUSSION OF POSSIBLE PHYSICAL MECHANISMS

The agreement obtained between the experimental results on time decrease of permeability and (a) the theoretical analysis above and (b) the experimental values obtained from magnetic viscosity measurements demonstrates the essential correctness of the activation concepts postulated to account for both phenomena. Any mechanism suggested must be compatible with the assumptions made in the development of the formal theory. One possible mechanism, already suggested in I, is based on the theory recently put forward by Stoner and Wohlfarth (1948) to account for the high values of the coercivity of permanent magnet alloys. These materials are supposed to consist of small ferromagnetic islands embodied in a less ferromagnetic matrix and the sizes of the islands are such as to preclude the formation of domain boundaries within them. It is shown that over certain ranges of the hysteresis cycles of such materials magnetization proceeds by discontinuous and irreversible rotations of the single domain vectors. In a simplified case, when the islands are prolate ellipsoids orientated at random, the discontinuous rotations occur over a range on the hysteresis cycle lying between a coercivity point and saturation. At any point in this range a certain number of domains have their magnetization vectors in positions of metastable equilibrium; that is, the free energy of a domain has a minimum value both before and after a discontinuous change of orientation, and transition from the initial to the final state may only occur if sufficient energy to overcome the intervening potential barrier is supplied. Thus Stoner's analysis would account for the potential barriers which may be overcome by thermal activation energy. In addition, as has already been shown in I, this theory would explain (a) the assumption that during the activation process no new domains of activation energy E are produced by successful activation processes, and (b) the experimentally observed fact that $\bar{t}p$ is independent of temperature.

Néel (1949 a) has examined the phenomenon of magnetic viscosity occurring in materials having a small quantity of ferromagnetic substance dispersed within them. He supposes that the ferromagnetic substance occurs in the form of particles of size below that necessary for domain boundary formation, i.e. single domains embedded in a non-ferromagnetic matrix are present. The postulate of thermal activation and the resulting $\log t$ relationship obtained by Néel are essentially similar to those advanced by the present authors and summarized in a note (Street and Woolley 1949 a). In addition, Néel has also suggested a mechanism whereby thermal deformation of the crystal lattice of the single domains may give rise to discontinuous rotations of the magnetization vectors

In a subsequent paper, Néel has considered the phenomenon of magnetic viscosity in 'massive' materials (1949b), and has shown that displacements of the domain boundary walls, due to thermal activation, may occur. This is a possible mechanism to account for magnetic viscosity phenomena in massive materials, and Néel states that the general results of this analysis are similar to those obtained in the consideration of single domain materials. Further information on this work has not yet been published, but, as far as the results on alnico are concerned, it is difficult to see how such irreversible boundary displacements are consistent with a value of $\bar{\eta}p$ independent of temperature and the non-production of new metastable states after successful activations. Further, any proposed mechanism should be capable of giving a satisfactory account of the high coercivity values of alnico, and in this respect the theory of Stoner and Wohlfarth appears to be the more likely.

§ 6. EFFECT OF HEAT TREATMENT

Measurements of magnetic viscosity and time decrease of permeability have been made using alnico specimens which have been subjected to various heat treatments. The conclusion reached is that, if the heat treatment is such as to produce material of highest coercivity, then the magnetic viscosity effects are substantially reduced. It is hoped to publish further details of this work at a later date.

ACKNOWLEDGMENTS

The authors thank Professor L. F. Bates for the facilities of his laboratory and his continued interest and encouragement in this work, and Dr. K. Hoselitz for supplying the specimen used.

REFERENCES

- GLAISHER, J. W. L., 1870, *Phil. Trans. Roy. Soc.*, **160**, 367.
NÉEL, L., 1949 a, *Ann. Géophys.*, **5**, 99; 1949 b, *C. R. Acad. Sci., Paris*, **228**, 1210.
SNOEK, J. L., 1938, *Physica*, **5**, 663; 1941, *Ibid.*, **8**, 711.
STONER, E. C., and WOHLFARTH, E. P., 1948, *Phil. Trans. Roy. Soc.*, **240**, 599.
STREET, R., and WOOLLEY, J. C., 1949 a, *Proc. Phys. Soc. B*, **62**, 141; 1949 b, *Proc. Phys. Soc. A*, **62**, 562; 1949 c, *Ibid.*, **62**, 743.
WEBB, C. E., and FORD, L. A., 1934, *J. Instn. Elect. Engrs.*, **75**, 787.

The Magnetic Properties of Uranium and Uranium-Iron Alloys

By L. F. BATES AND J. R. MALLARD

University of Nottingham

MS. received 6th January 1950

ABSTRACT. Measurements have been made of the magnetic susceptibility of fairly pure uranium and of 5 and 10 atomic per cent alloys of iron with uranium. Using a modification of the Curie method, with special pole-pieces and electromagnetic dynamometer control, the specimens were found to be paramagnetic with no trace of ferromagnetism in the temperature range 20° to 350° C. At 20° C. the susceptibility of pure uranium is $+1.740 \times 10^{-6}$ E.M.U. per gm., increasing to $+1.804 \times 10^{-6}$ at 350° C.; the results fit the equation

$$\chi_U = 32 \times 10^{-11} T + 1.564 \times 10^{-6} + 24.0 \times 10^{-6} / T.$$

Assuming that the iron is present in form of the compound U_6Fe , it is found that the latter has a susceptibility of $+2.036 \times 10^{-6}$ E.M.U. per gm. at 20° C., which remains practically constant with temperature.

§ 1. INTRODUCTION

IN the last of a series of papers on the magnetic properties of metals dissolved in mercury, referred to hereinafter as A, Bates and Somekh (1944) described an apparatus for measuring the magnetic susceptibility of non-homogeneous systems and its use in studying cadmium, gold, silver and zinc amalgams over wide ranges of concentrations at room temperature. In this work ferromagnetic impurities were rigorously avoided, so that little or no variations of field conditions were required. As the Curie method was used, the amalgams were enclosed in one of two long cylindrical tubes mounted so that they could move along magnetic lines of force running between special cylindrical pole pieces in such a way that the force exerted on each element of amalgam was as nearly as possible independent of its position in the tube (cf. Bates 1948). Details of the main features of the apparatus and the basic theory of measurement by a dynamometer method are given in A, pp. 183 to 185.

It was decided, in view of its present-day importance, to measure the susceptibility of uranium, kindly supplied by the Atomic Energy Research Establishment at Harwell, with the above apparatus improved in certain particulars as described below. It was clearly necessary to deal with possible effects of ferromagnetic impurities and also to make measurements above room temperature. The specimen tube was therefore mounted inside a vertical, cylindrical electric furnace consisting of a copper tube wrapped with asbestos paper, wound non-inductively with nichrome strip, and coated with 'purimachos' cement. The whole was placed inside a brass cooling jacket through which water was continuously circulated. A copper-constantan thermocouple with one junction placed directly below the specimen tube indicated the temperature; the leads to the junction passed through holes drilled in a 'sindanyo' base to which the jacket was fixed.

In A, the specimen tubes were of such a length that the centre of the specimen was located some 18 cm. below the glass hook of the stopper, the latter being hooked on to the copper framework extension of the electro-dynamometer with a touch of

paraffin wax melted upon the junction. When the furnace temperatures rose above 100 to 150° c. this wax joint melted and the tube sometimes changed its position. Accordingly, in both the glass hook and the copper hook a well defined V notch was filed, and a small length of plaited silk, with a loop at each end, was used to connect the hooks and suspend the tube. Lateral displacements due to temperature changes were thus effectively eliminated and the tube could hang vertically with the stopper in a standard position, whether empty or containing liquid or solid specimens which fitted the tube. Of course, solids which did not exactly fit the tube might displace it from the vertical and so alter the effective length of the couple arm of the electro-dynamometer. However, by repeating the measurements in such a case with the tube carefully rotated through 180° so that the displacement is reversed, this defect could, clearly, be overcome. The silk connection also had the unexpected advantage that accidental collisions of the tube with the furnace walls did not cause trouble.

The apparatus was calibrated with mercury distilled by the Hulett method in an all-glass apparatus to reduce contamination; the susceptibility of mercury was taken as -0.1680×10^{-6} E.M.U. per gm. The specimen tubes were evacuated by the simple method described in A. A new electro-dynamometer was constructed, with the coil wound on a former in order to give less distortion and to deal with larger couples.

The A.C. current to the dynamometer coils was supplied by a constant voltage transformer delivering 230 volts r.m.s. at 500 watts, the voltage fluctuations being less than 0.2%. The same electromagnet was used as in A, but it was supplied from a battery of heavy-duty accumulators.

§ 2. EXPERIMENTAL MEASUREMENTS

For preliminary tests a specimen of pure uranium was turned down from a small rod, taking the recommended health precautions during the turning and handling, to form a cylinder 1.5 cm. long and 5 mm. diameter weighing about 5 gm. To prevent rapid oxidation of the freshly turned surface the uranium was dropped into a dish of concentrated nitric acid—which dissolves the oxide but only attacks the metal slowly—quickly transferred to a dish of absolute alcohol and washed thoroughly in the latter. After rapid drying with filter paper it was slipped into the specimen tube, which was then evacuated rapidly. Less speed in manipulation was necessary with iron-uranium alloys which oxidize much more slowly.

The pure uranium was tested for ferromagnetism by measuring the susceptibility k_H for different fields H and plotting k_H as a function of $1/H$ (cf. Bates 1948, p. 114); if no impurity is behaving as a ferromagnetic substance the graph should be parallel to the $1/H$ axis. The values of H were small enough to ensure that the pole-pieces were far from saturated, so that the magnetic field was always of the form $H(dH/dr) = \text{const.}$ Special tests were also made to ensure that for a given value of H equal portions of the specimen were all exposed to the same magnetic force. This was done by measuring the forces on different volumes of liquid in the specimen tube and showing how the force depended on the volume occupied. As a final result of these tests, the specimen tube was never occupied to a height greater than 3 cm.; the occupied portion was kept as closely as possible midway between the upper and lower surfaces of the pole-pieces, and a field current of 3.6 amp. was not exceeded. This reduction in specimen length was compensated by using new specimen tubes of larger internal diameter.

For final measurements on pure uranium at room temperature a rod 1.35 cm. long closely fitting the specimen tube was used, and the restoring couples were found for some six suitable values of the field current. For a given value of the field current, the two restoring couples, with the specimen tube in a chosen position and in another turned through 180° with respect to the first, differed by less than 0.5%. Couples on a specimen of pure mercury were measured in the same way; so, too, were those on the empty specimen tube.

Let χ_s and χ_{Hg} be the mass susceptibilities of the specimen and Hg respectively; i_s , i_{Hg} and i_0 the dynamometer restoring currents for the tube containing the specimen of mass m_s , mercury of mass m_{Hg} , and vacuum, respectively, in the same field; then

$$\chi_s/\chi_{Hg} = (m_{Hg}/m_s)[(i_s^2 - i_0^2)/(i_{Hg}^2 - i_0^2)] = (m_{Hg}/m_s)[I_s^2/I_{Hg}^2].$$

A graph of I_U^2 ($I_U^2 = I_s^2$, but, strictly speaking, contains a small correction for twist in the suspension), plotted against I_{Hg}^2 is shown in Figure 1, curve (a).

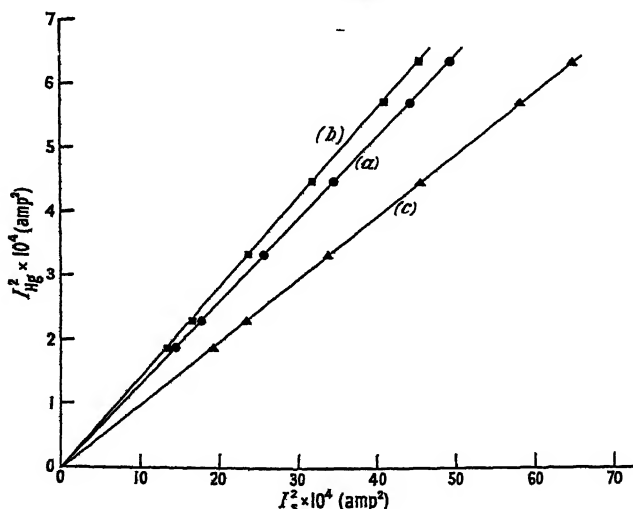


Figure 1. Test for absence of ferromagnetism.

It is undoubtedly linear, so that both the couple on the uranium specimen and that on the mercury were strictly proportional to $H(dH/dr)$. Consequently, the uranium could not have contained a ferromagnetic impurity acting as such. It is easy to provide a proof that ferromagnetic impurities in both the specimen and mercury could not give the line (a) of Figure 1.

Applying the Gaussian theory of errors to line (a) the value of the slope I_{Hg}^2/I_s^2 was found, whence the final result for the susceptibility of pure uranium was $1.740 \pm 0.003 \times 10^{-6}$ E.M.U. per gm. The error ± 0.003 is the probable error to be attached to the relative measurement; the maximum error is thus ± 0.015 and the standard deviation ± 0.005 .

After these measurements had been completed the specimen was analysed and found to contain 700 parts in 10^6 of iron, 60 in 10^6 of nickel and 5 in 10^6 of cobalt. Consequently, uranium behaves like aluminium and copper in that it may dissolve iron in such a way that the latter cannot exhibit ferromagnetism. The presence of so much iron in a ferromagnetic state would have given couples some twenty times as large as those observed above.

In view of this unexpected behaviour, steps were taken to acquire specimens of nominal iron contents of 5 and 10 atomic per cent. The 5% specimen was too small to fit exactly into the specimen tube so it was fitted with a cylindrical sheath of 'Perspex', for which due allowance was made in the measurements. The results are shown in curve (b) of Figure 1; curve (c) gives the results for the 10% specimen. Using the same theoretical procedure as before, the susceptibilities of the specimens were found to be $+1.842 \pm 0.003$ and $1.924 \pm 0.004 \times 10^{-6}$ E.M.U. per gm. respectively; and, even with as much as 10 atomic per cent impurity of iron, there was still no trace of ferromagnetism.

The variations of the susceptibilities of the three materials with temperature, measured as described, are shown in Figure 2. In some of these measurements

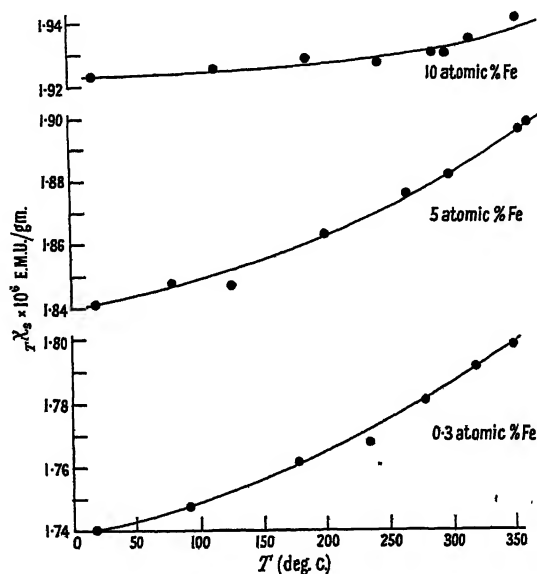


Figure 2. Variation of susceptibility of uranium-iron specimens with temperature.

convection currents caused some small displacement of the zero of the suspended system. Allowance for this was adequate at lower temperatures, but appeared uncertain above 350° C., and in view of the relatively small changes of susceptibility with temperature, measurements were not recorded above 350° C.

§ 3. DISCUSSION OF RESULTS

It is understood that uranium forms alloys U_6Fe and U_6Fe_2 with iron. In alloys of low iron content nearly all the iron would be expected to be in combination as U_6Fe . Assuming no interaction between uranium metal and U_6Fe the susceptibility of an iron-uranium alloy should be the sum of the susceptibilities of pure uranium and the compound U_6Fe according to Wiedemann's law; so that, if a specimen contains x atomic per cent of iron, there will be present a mass $(xM_{Fe} + 6xM_U)$ of U_6Fe , where M_{Fe} and M_U are the atomic weights of iron and uranium respectively. Hence, the susceptibility χ_s of such a preparation should be given by

$$\chi_s[xM_{Fe} + (100 - x)M_U] = \chi_{U_6Fe}(xM_{Fe} + 6xM_U) + \chi_U(100 - 7x)M_U. \quad \dots\dots(1)$$

Now the correct iron contents for the three materials as established by later analysis were 0.2978, 4.884 and 9.491 atomic per cent, respectively. Consequently, we have the three equations:

$$23,760(1.740 \pm 0.003)10^{-6} = 442.1\chi_{\text{U}_6\text{Fe}} + 23,320\chi_{\text{U}}, \quad \dots\dots(2)$$

$$23,930(1.842 \pm 0.003)10^{-6} = 7252\chi_{\text{U}_6\text{Fe}} + 15,670\chi_{\text{U}}, \quad \dots\dots(3)$$

$$22,080(1.924 \pm 0.004)10^{-6} = 14,090\chi_{\text{U}_6\text{Fe}} + 7,991\chi_{\text{U}}, \quad \dots\dots(4)$$

These three equations containing two unknowns will not give the best values of χ_{U} and $\chi_{\text{U}_6\text{Fe}}$ if taken in pairs and treated simultaneously, as each result so obtained does not take account of all the experimental data. As there are only three equations, we used, not the method of least squares, but the more simple 'least sum' method of computation. Half of equation (3) was added to each of equations (2) and (4) to give two 'normal' equations which were solved simultaneously for χ_{U} and $\chi_{\text{U}_6\text{Fe}}$, giving $\chi_{\text{U}} = +1.740 \pm 0.003 \times 10^{-6}$ and $\chi_{\text{U}_6\text{Fe}} = +2.036 \pm 0.007 \times 10^{-6}$ E.M.U. per gm.

By taking values from all three curves of Figure 2 at any chosen temperature, then χ_{U} and $\chi_{\text{U}_6\text{Fe}}$ could be obtained at that temperature by calculations exactly similar to those for room temperature just described. Figure 3 (a) shows the

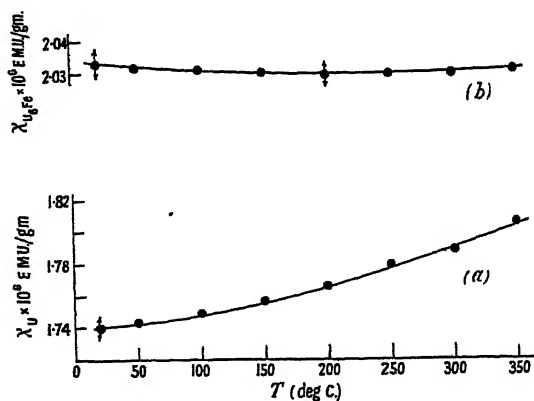


Figure 3. Variation of susceptibilities of U and U_6Fe with temperature.

curve obtained for uranium metal, and it shows that there is a steady increase of susceptibility with temperature of the order of 3.5 per cent over the range 20–350°C. The corresponding curve for $\chi_{\text{U}_6\text{Fe}}$ is given in Figure 3 (b), which apparently shows a very small temperature decrease of $\chi_{\text{U}_6\text{Fe}}$ over the same range. The arrows at 20°C. indicate the probable error in this part of the work, viz. ± 0.007 , whilst those at 200°C. show the probable error in any of the other temperature measurements, viz. ± 0.008 . This error is larger, because of the errors to be associated with the high temperature value relative to the room temperature value. It is seen that the error range indicated shows that no great reliance can be placed on the small decrease of $\chi_{\text{U}_6\text{Fe}}$ with temperature, i.e. within probable experimental error $\chi_{\text{U}_6\text{Fe}}$ is independent of temperature in the range 20–350°C.

It must be pointed out that these results give no real evidence to support the assumption that the iron actually exists as U_6Fe . Now, the equation (1) could be modified because each specimen might be supposed to contain, for

example, uranium and some non-ferromagnetic form of iron not combined with uranium as a compound, or even to contain uranium and UFe_2 , which is also known to exist. It was therefore thought that it might be possible, by considering the final error to be attached to χ_U as derived from these various forms of equation (1), to obtain an indication as to which form of equation (1) is the most correct. Calculations were carried out using the relation

$$\chi_s(xM_{\text{Fe}} + (100-x)M_U) = \chi_{\text{Fe}}xM_{\text{Fe}} + \chi_U(100-x)M_U,$$

i.e. assuming only uranium and iron in a paramagnetic form to be present. The value obtained for χ_U was very close to the one quoted above, and the error to be attached to the value was still of the order of ± 0.003 , thus rendering the calculations inconclusive.

No account has yet been taken of other ferromagnetic elements present as impurity, viz. cobalt and nickel. Cobalt can be completely neglected since all specimens contained less than 20 parts in 10^6 . Nickel, however, was present in about 95 parts in 10^6 and, assuming that nickel was present as U_6Ni , equation (1) was written in the form

$$\chi_s[xM_{\text{Fe}} + zM_{\text{Ni}} + (100-x-z)M_U] \\ = \chi_{\text{U}_6\text{Fe}}(xM_{\text{Fe}} + 6zM_U) + \chi_{\text{U}_6\text{Ni}}(zM_{\text{Ni}} + 6zM_U) + \chi_U(100-7x-7z)M_U,$$

where z is the atomic per cent of nickel present. From the three equations so obtained, one for each specimen, values of χ_U , $\chi_{\text{U}_6\text{Fe}}$ and $\chi_{\text{U}_6\text{Ni}}$ could be obtained, and the calculations showed that, as expected, the values of χ_U and $\chi_{\text{U}_6\text{Fe}}$ were not appreciably changed. The error to be attached to $\chi_{\text{U}_6\text{Ni}}$ was, however, so large that the result was valueless.

It is possible to make some deductions concerning the behaviour of pure uranium by noting that the experimental results of Figure 3 fit the equation

$$\chi_U = 32 \times 10^{-11}T + 1.564 \times 10^{-6} + 24.0 \times 10^{-6}/T,$$

since this shows the great preponderance of a temperature independent term in the expression for the susceptibility. If we isolate the last term in the above equation, we find that it is equivalent to a magnetic carrier of $0.21 \mu_B$ obeying the Curie law. It is interesting that Pauling (1947, see also Zachariasen 1949), suggested the allocation to uranium of 5.78 instead of 6 valency electrons, from which we now deduce a possible magneton number of 0.22 for the electron residue; the agreement is somewhat striking.

Suppose, however, that in uranium metal we postulate that the six valency electrons are free and that we allow -24.3×10^{-6} E.M.U. per gm. atom for the radon core, then the gramme atomic paramagnetism at room temperature is 438.6×10^{-6} . Applying Stoner's (1935, 1936) free electron theory we find that the theoretical susceptibility should be $32.1q/V_0 \times 10^{-6}$ where q is the number of free electrons per atom and V_0 is the width of the energy band in electron volts. The experimental data give $V_0 = 0.4392$ compared with the value 15.52 for completely free electrons. Allowing for the constant term in the preceding paragraph the value would be approximately 4.6. The calculated, $(\Delta\chi_U/\chi_U \cdot \Delta T)$, temperature decrease of susceptibility, assuming that the characteristic temperature of uranium is 4220°K. , is -0.29×10^{-4} per degree c.; this is opposite in sign to the observed effect, but the discrepancy between experiment and theory may to some extent be attributed to the thermal expansion of the metal. Indeed, the values of the coefficient of linear expansion, namely 16×10^{-6} and 23×10^{-6} per degree c. in the temperature ranges 20 to 100°C. and 300 to 400°C. , respectively, would

lead us to expect a small increase of susceptibility with temperature, although it would certainly be much smaller than the experimental value.

It is not thought necessary to compare our results with those of Owen (1912) and Honda (1910) because these workers are unlikely to have used reasonably pure uranium. Since this work was completed the paper by Bommer (1941) has come to hand. He prepared a mixture of uranium and potassium chloride by heating UCl_4 with excess metallic potassium, later distilling off the excess potassium. Bommer's value of χ_U at room temperature is 620×10^{-6} E.M.U. per gm. atom, and he finds it to decrease slightly with increase in temperature. However, as one of his specimens showed ferromagnetism at liquid air temperatures, while two others did not, we feel that there is no point in discussing his results in detail. To carry out such low temperature measurements a new apparatus is necessary, and this is now being made.

Perhaps a few words are pertinent here concerning the evidence for the existence of 5 f electrons in uranium; much of it has recently been summarized by Meggers (1947). We have Mayer's (1941) theoretical treatment of the potential functions of 4 f and 5 f electrons, from which he concluded that a series of 'transuranic elements' should commence filling 5 f shells at $Z=90$ to 92. We have the chemical evidence of Seaborg (1946), Emeléus (1947) and Starke (1947) indicating that the first 5 f electron appears at $Z=90$. Then there is the spectroscopic evidence of Kiess, Humphreys and Laun (1946) who conclude from the spectrum of uranium that three 5 f electrons are present. The heavier metallic elements which show temperature increase of paramagnetism are Ru, Rh and Ir, belonging to the palladium and platinum groups, and it may be significant that in all cases the ground term is usually considered as an F state.

To sum up, the magnitude of the susceptibility and the absence of a Curie or Curie-Weiss temperature relation point to the conclusion that uranium must be treated as a transition element.

ACKNOWLEDGMENTS

We record our thanks to the Director and to Mr. B. W. Mott of the Atomic Energy Research Establishment, Harwell, for the provision of samples of uranium and uranium-iron alloys and for their interest in this work. One of us (J. R. M.) thanks the Department of Scientific and Industrial Research for a maintenance allowance.

REFERENCES

- BATES, L. F., 1948, *Modern Magnetism* (Cambridge : University Press).
 BATES, L. F., and SOMEKH, E. M., 1944, *Proc. Phys. Soc.*, **56**, 182.
 BOMMER, H., 1941, *Z. Anorg. u. Allgem. Chemie*, **247**, 249.
 EMELÉUS, H. J., 1947, *Some Aspects of Nuclear Chemistry* (The Royal Institute of Chemistry).
 HONDA, K., 1910, *Ann. Phys., Lpz.*, **32**, 1044.
 KIESS, C. C., HUMPHREYS, C. J., and LAUN, D. D., 1946, *J. Res. Nat. Bur. Stand., Wash.*, **57**, 57.
 MAYER, M. G., 1941, *Phys. Rev.*, **60**, 184.
 MEGGERS, W. F., 1947, *Science*, **105**, 514.
 OWEN, M., 1912, *Ann. Phys., Lpz.*, **37**, 657.
 PAULING, L. C., 1947, *J. Amer. Chem. Soc.*, **69**, 542.
 SEABORG, G. T., 1946, *Science*, **104**, 379.
 STARKE, K., 1947, *Naturwissenschaften*, **69**, 34.
 STONER, E. C., 1935, *Proc. Roy. Soc. A*, **152**, 672; 1936, *Ibid.*, **154**, 656.
 ZACHARIASEN, W. H., 1949, *Acta Crystallogr.*, **2**, 94.

On Phase-Contrast Testing with a Slit Source

By E. H. LINFOOT

The Observatory, Cambridge

MS. received 10th November 1949

ABSTRACT. In the standard form of the phase-contrast test, the light-source is a pinhole comparable in size with the Airy disc of the system under test. As a consequence, it is difficult to obtain enough light through the pinhole to carry out the test with full efficiency. If a slit source and a phase-changing strip are used this difficulty is overcome, but the properties of the test are altered; for example, the brightness distribution seen on the surface of a true mirror under the test is no longer radially symmetrical.

In the present paper the properties of the phase-strip test are examined in the mathematically simplest case where the light-source is a slit of negligible width, passing monochromatic light, and it is concluded that the test can safely be used in figuring an optical surface to an accuracy of about one twentieth of a fringe. Below this limit, the systematic error inherent in the usual interpretation of the test begins to be appreciable in the case of slow errors, though rapid local errors are correctly seen right up to the limit of sensitivity of the test.

§ 1. INTRODUCTION

THE elegant and ingenious phase-contrast method of testing the figures of optical surfaces, devised about sixteen years ago by F. Zernike (1934 a), has been described from the experimental side by C. R. Burch (1934) and from the theoretical by its originator and by the present writer (Zernike 1934 b, Linfoot 1946 a). Dr. Burch has described the setting up of the test essentially as follows:

An artificial star of small diameter is placed at one of the stigmatic points of an optical system and at the other stigmatic point there is placed the small circular phase-retarding disc or Zernike disc. The star and the Zernike disc are both preferably somewhat smaller than the Airy disc of the system at the points at which they are respectively placed. The eye of the observer is placed close behind the Zernike disc, and he looks either directly at the surface under test or at its image in any auxiliary surfaces which may come between it and him. He then sees interference fringes formed between the wave leaving the surface and the supplementary wave to which its passage through the Zernike disc gives rise. If the Zernike disc is not larger than the Airy disc produced by the system, that part of the supplementary wave lying within the cone of light proceeding from the system will be substantially spherical, so that (since the wave front leaving the system departs from sphericity only by the errors of the mirror under test) the fringe system seen in a given wavelength will be substantially the Newton-ring system which would be seen if a true-figure test plate were laid on, or very close to, the surface under test. The colour sequence of the fringes will be determined by the chromatic difference of phase-retardation of the Zernike disc, and can be ascertained from the colour changes which occur when the disc is moved out of focus, and when it is decentred.

A practical drawback to this form of the test is the difficulty of getting enough light through the very small pinhole which has to be used. A high pressure

mercury arc used with a good-quality condenser will overcome the difficulty to some extent, but even so the test needs to be carried out in a darkened room, and in testing an unsilvered glass mirror the amount of light is inconveniently small.

In the Foucault knife-edge test, this shortage of light can be overcome very simply by substituting a fine slit for the pinhole and testing with the knife-edge set as accurately parallel as possible to the image of the slit. The forms and relative intensities of the shadows seen under the test are not altered by the change, provided only that the slit is too short for the aberrations of the system to vary appreciably along its length, and the amount of light available for testing purposes is increased by a factor of 50 or 100.

It was pointed out by Zernike that a slit-source can also be used in the phase-contrast test provided that the phase-changing disc is replaced by a phase-changing strip, adjusted to cover the diffraction image of the slit. Strips of suitable dimensions can be made without very much difficulty by pressing fine threads of resin between glass, or by stroking the surface of a glass plate with a 'cutanit' (molybdenum-titanium carbide) tool*, and fairly rapid irregular errors are found to show very nearly the same fringe-patterns under test with a strip as with a disc.

Thus the phase-strip test shares one of the most valuable properties of the Zernike-disc test, that of displaying rapid irregular errors of figure on the mirror surface in a way which allows their size and shape to be correctly 'seen' by simple inspection of the surface under test.

Of even greater importance to the practical mirror-maker, however, is the question whether the strip test can safely be used for the control of the three errors which are most commonly encountered during the figuring of an optical surface, namely primary spherical aberration, zonal error and astigmatism. The examination of this question is the main object of the present paper.

In §2 formulae are derived which predict the intensities seen under the test on a surface with arbitrary smooth errors. In §3, these formulae are applied to the case of a nearly true mirror and it is shown that something near to the best results may be expected with a quarter-wave strip, of width about one third of the diameter of the Airy disc. The appearance, under test with such a strip, of small amounts of primary astigmatism, of primary spherical aberration, and of local zonal error is discussed in §4, and from the examination of special cases it is inferred that the test can safely be used for the figuring of optical surfaces to an accuracy of about one twentieth of a fringe.

Throughout the analysis it is assumed that the absorption and back-reflection of the strip may be disregarded, that the width of the slit is small compared with the Airy disc of the system under test and that the light is monochromatic. In practice, the width of the slit would usually be comparable in size with the diameter of the Airy disc and the light polychromatic, but a discussion of the extent to which the properties of the test are modified by these circumstances is beyond the scope of the present paper.

§ 2. DIFFRACTION THEORY OF THE PHASE-CONTRAST STRIP TEST

We consider first the case where the light originates in a point source, the geometrical image of which falls on the centre of the phase-changing strip. The notation and assumptions are similar to those used in earlier papers (Linfoot 1946 a, b). For convenience they are given again here.

* C. R. Burch (1934). The engraved strips advance the phase, giving 'positive phase contrast'.

All the cases when a convergent pencil is being null-tested can be covered by supposing that the wave fronts originate at the surface of a nearly spherical mirror M which is being tested at its approximate centre of curvature. The results of the test are then interpreted in terms of errors of figure of the mirror M . We suppose that these errors of figure may amount to several wavelengths, but that the error slopes on the mirror are not so steep as to spread out the visible image to more than a moderate multiple, say 5 or 10, of the size of the Airy disc.

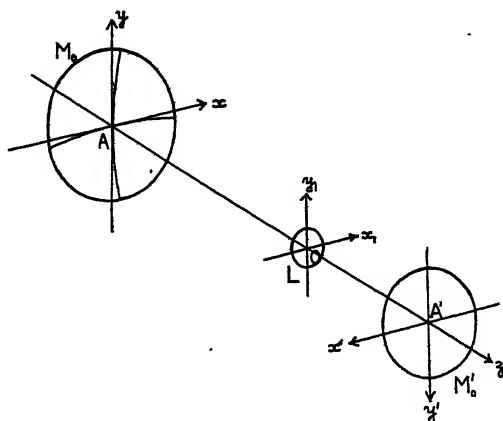


Figure 1.

Then it makes no appreciable difference if we suppose the wave to be leaving a true spherical surface M_0 (Figure 1), lying everywhere within a few wavelengths of the surface M , and suppose that at the point (x, y, z) on M_0 the complex displacement is

$$E(x, y) = |E(x, y)| \exp \{ (2\pi i / \lambda) \phi(x, y) \};$$

thus $|E(x, y)|$ is the amplitude at the point (x, y) on M , and $\phi(x, y)$ the phase there, while λ is the wavelength of the light. The variation in $\phi(x, y)$ expresses the distortion of the wave fronts. Outside the boundary-circle C , lying in M , and corresponding to the edge of the mirror, $E(x, y)$ is defined as zero.

After leaving the mirror, the light comes to a more or less imperfect focus in the neighbourhood of the centre of curvature O of M , at a distance s from its pole A ; we suppose that s is large compared with the diameter of the mirror. Just behind O , a viewing system L receives the light and forms and image M'_0 of the surface M ; we denote by C' the boundary of M'_0 .

At O is placed the phase-changing strip which renders variations of phase in the wave fronts leaving M_0 visible as variations of intensity in the image-surface M'_0 ; x_1 and y_1 , measured parallel to x and y , are coordinates in the space near O . The central line of the strip lies along the axis Oy_1 in the plane T drawn through O at right angles to Oz .

We define a system of coordinate-numbers (x', y') in the surface M'_0 , by assigning to the image P' of the point $P(x, y, z)$ on M_0 the coordinate-numbers $x' = x, y' = y$. The intensity at P' is thus the same thing as the 'intensity seen under the test' at P . To avoid unessential complications, we suppose that L is a perfect optical system.

By the usual arguments, based on an application of Huyghens' principle, the complex displacement in the intermediate image surface, which we take to be the sphere S of centre A passing through O , is

$$W(u, v) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{ixu + iyv} E(x, y) dx dy, \quad \dots\dots (2.1)$$

where

$$u = \frac{2\pi x_1}{\lambda s}, \quad v = \frac{2\pi y_1}{\lambda s}, \quad \dots\dots (2.2)$$

and the integral is only formally over an infinite domain, since the integrand vanishes outside the boundary contour C .

From the assumptions made concerning the form of the errors of the surface under test, it follows that almost all the light crosses the intermediate image surface in the immediate neighbourhood of the point O .

If the phase-changing strip were not present, the complex displacement $D(x', y')$ at the point (x', y') in the final image-surface would be obtained by integrating the product

$$\frac{1}{2\pi} \exp \{ -i(x'u + y'v) \} \cdot W(u, v)$$

over the area $u^2 + v^2 \leq R^2$ of the u, v plane which falls within the aperture of the viewing system L . (In the case of an $f/10$ pencil and a viewing telescope of 1 inch aperture, the value of R is approximately 1,300.) The effect of an idealized Zernike strip occupying the region $-c \leq u' \leq c$ of the intermediate image surface and retarding by α radians the phase of the waves passing through it can be represented by changing $W(u, v)$ into $e^{-i\alpha} W(u, v)$ throughout this region and leaving it unaltered elsewhere. In saying this, we assume that the effect of the strip is merely to change the phase of those parts of the wave which impinge upon it, without influencing the neighbouring parts. This assumption is never strictly true, but it is a permissible approximation when the pencil under test is of small numerical aperture.

In reality, the phase-strip is not an infinitesimally thin lamina but a solid body of length $10,000\lambda$ or more, of width 10 to 100λ , and of thickness comparable with λ . Now throughout that u, v region of S in which an alteration in the values of $W(u, v)$ can cause a visible change in the appearance of the mirror under the test, the distance between the surface S and the plane T is of the order of $10^{-4}\lambda$ in a practical case. For example, when a 12-inch mirror is tested at $f/10$ in light of wavelength $\lambda = 2 \times 10^{-5}$ inch, the distance between S and T is less than $5 \times 10^{-5}\lambda$ throughout the circle $u^2 + v^2 \leq 14.7 \times 10^4$ of radius 100 times that of the Airy disc, while the replacement of $W(u, v)$ by zero outside this circle does not sensibly affect the appearances seen under the test.* It follows that, to the order of approximation here involved, the idealized phase-changing lamina which replaces our three-dimensional phase-strip in the analysis may as well be supposed to lie

* This may be verified experimentally by testing a nearly true mirror of long focal ratio (about $f/50$) with an iris diaphragm placed in front of the viewing telescope. (See Linfoot (1946 b) for an account of the corresponding experiment on changes in the appearances seen under the Foucault knife-edge test.)

in the intermediate image surface S as in the surface T . The complex displacement at the point (x', y') of the final image surface is therefore given, to a sufficient approximation, by the equation

$$D(x', y') = \frac{1}{2\pi} \iint_{u^2 + v^2 \leq R^2} \exp(-iux' - ivy') \cdot W(u, v) du dv \\ + \frac{1}{2\pi} (e^{-i\alpha} - 1) \iint_{u^2 + v^2 \leq R^2, -c \leq u \leq c} \exp(-iux' - ivy') \cdot W(u, v) du dv. \quad \dots\dots(2.3)$$

The first term of (2.3) is simply the complex displacement which results when the mirror is imaged through the viewing system, without the interposition of the Zernike strip, and the second term therefore describes the effect of the strip.

We next replace R by ∞ in (2.3). This amounts to neglecting the effects of the finite aperture of the viewing telescope. For mirrors with errors of the type under consideration, the resulting change in the predicted intensities is too small to be detected by the eye, except at points whose distance from the rim of the mirror is comparable with the resolving power of the viewing system. If, in order to cover the possibility of a central obstruction or of a sharp step on the surface of M , the hypotheses are widened by allowing $E(x, y)$ to be discontinuous along a finite set of arcs of differentiable curves, the above assertion remains valid provided we replace the words "rim of the mirror" by "set of points of discontinuity of $E(x, y)$ ". At these points the intensities depend critically on the aperture and optical aberrations of the viewing system, as well as on the errors of the mirror, and we therefore exclude them from the discussion. At the remaining points we obtain, on setting $R = \infty$ in (2.3), the equation

$$D(x', y') = \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \exp(-iux' - ivy') \cdot W(u, v) du dv \\ + \frac{1}{2\pi} (e^{-i\alpha} - 1) \iint_{-c \leq u \leq c, -\infty < v < \infty} \exp(-iux' - ivy') \cdot W(u, v) du dv. \quad \dots\dots(2.4)$$

On substituting for $W(u, v)$ from (2.1) and applying Fourier's inversion formula, (2.4) becomes

$$D(x', y') = E(x', y') + \frac{1}{\pi} (e^{-i\alpha} - 1) \int_{-c}^c e^{-iux'} du \int_{-\infty}^{\infty} e^{iux} E(x, y') dx \\ = E(x', y') + \frac{1}{\pi} (e^{-i\alpha} - 1) \int_{-\infty}^{\infty} E(x, y') \frac{\sin c(x - x')}{x - x'} dx. \quad \dots\dots(2.5)$$

These transformations are justified because $E(x, y)$ is zero everywhere outside a finite region of the x, y plane, is of bounded variation in this region, and is differentiable at the point $x = x', y = y'$.

The intensity seen under the test, at the point (x', y') on the mirror-surface or in the halo, is then measured by the quantity

$$I(x', y') = |D(x', y')|^2. \quad \dots\dots(2.6)$$

So far, the light-source has been assumed to be a pinhole of negligible dimensions. In the case of a slit of negligible width, each point of the slit can be regarded as a point source. If each point of the primary light source were sharply focused

on to the plane of the slit by a wide-aperture condenser system of negligible aberrations, the different points of the slit could be treated as independent sources. In such a case, the observed intensity under the test would simply be the sum of the intensity contributions from the different point sources. If we suppose that the length of the slit is not great enough for off-axis aberrations of the system to make an appreciable contribution to the apparent errors of the surface under test, the intensity distributions contributed by any two points of the slit differ only by a constant factor. Therefore, the intensity function $I(x', y')$ is still given by equation (2.6).

However, in a practical case the aperture of the condenser system is often comparable with that of the system under test and it is not true that all points of the primary source are sharply focused on to the plane of the slit by the condenser. Each luminous particle W of the primary source contributes a wave displacement

$$E_W(x, y) = A_W(x, y)E(x, y)$$

on the surface of M_0 , where $E(x, y) = \exp \{ (2\pi i / \lambda) \phi(x, y) \}$ inside C and zero outside C , as before, while $A_W(x, y)$ expresses the characteristics of the wave emitted by the W particle when, after passing through the condensing system and the slit, it arrives at the point (x, y) on the nearly spherical mirror M . Because the width of the slit is negligible, $A_W(x, y)$ does not vary appreciably with x , and we may write, to a sufficient approximation

$$E_W(x, y) = A_W(y)E(x, y). \quad \dots\dots(2.7)$$

$|E_W(x, y)|^2$ measures the intensity of the wave at this point, so that if the mirror is to appear uniformly lit in the absence of the phase strip

$$\sum_W |E_W(x, y)|^2$$

must be constant over its surface. By a suitable choice of units we can write this condition in the form

$$\sum_W |A_W(y)|^2 = 1 \quad \text{on } M_0. \quad \dots\dots(2.8)$$

The complex displacement contribution from the W particle in the final image surface is then obtained on replacing $E(x, y)$ by $A_W(y)E(x, y)$ in the previous analysis; it is

$$\begin{aligned} D_W(x', y') &= A_W(y')E(x', y') + \frac{1}{2\pi} (e^{-i\alpha} - 1) \int_{-\infty}^{\infty} A_W(y')E(x, y') \frac{\sin c(x - x')}{x - x'} dx \\ &= A_W(y')D(x', y'), \end{aligned}$$

where $D(x', y')$ is given by (2.5). Since the waves from the different W particles are mutually incoherent, the total intensity seen on M_0 under the test is now

$$\begin{aligned} \sum_W |D_W(x', y')|^2 &= \sum_W |A_W(y')|^2 |D(x', y')|^2 \\ &= |D(x', y')|^2, \end{aligned}$$

by (2.8). That is, the intensities seen on M_0 are still given by (2.6). The same is true of the intensities in the halo, since these differ from zero only in the y -range where (2.8) holds.

From (2.5) we see that the value of $D(x', y')$ at any point (x', y') depends only on the figure of the mirror M along the horizontal line $y = y'$ and that the halo is confined to a horizontal band filling the same y -range as do the points of the mirror. In these two respects the phase-strip test resembles the knife edge test. On the other hand, since the function $\sin \{c(x - x')\}/(x - x')$ is equivalent to a function of x everywhere continuous, it can be shown that there is no bright 'Rayleigh ring' at the edge of the mirror, but an apparent ordinary discontinuity in $I(x', y')$ arising from the first term on the right of (2.5). The same is true at a step-discontinuity on the surface of M .

§ 3. NEARLY TRUE MIRROR; CHOICE OF PHASE STRIP

On a nearly true mirror of uniform reflecting power, we can write

$$\begin{aligned} E(x, y) &= 1 + iE^*(x, y) & (x^2 + y^2 \leq H^2) \\ &= 0 & (x^2 + y^2 > H^2), \end{aligned} \quad \dots\dots (3.1)$$

where $E^*(x, y)$ is real and so small that its square can be neglected in comparison with unity, while $2H$ is the diameter of the mirror surface.* (2.5) then gives, for $x^2 + y^2 \leq H^2$,

$$\begin{aligned} D(x, y) &= 1 + iE^*(x, y) - [(1 - \cos \alpha) + i \sin \alpha] \frac{1}{\pi} \int_{-Y}^Y (1 + iE^*(t, y)) \frac{\sin c(t - x)}{t - x} dt, \\ &= 1 - (1 - \cos \alpha) I_c(Y) + \sin \alpha \frac{1}{\pi} \int_{-Y}^Y E^*(t, y) \frac{\sin c(t - x)}{t - x} dt \\ &\quad + i \left[E^*(x, y) - \sin \alpha I_c(Y) - (1 - \cos \alpha) \frac{1}{\pi} \int_{-Y}^Y E^*(t, y) \frac{\sin c(t - x)}{t - x} dt \right], \end{aligned} \quad \dots\dots (3.2)$$

where Y is written for $(H^2 - x^2)^{1/2}$ and

$$I_c(Y) = \frac{1}{\pi} \int_{-Y}^Y \frac{\sin c(t - x)}{t - x} dt = \frac{1}{\pi} [\text{Si}\{c(Y - x)\} + \text{Si}\{c(Y + x)\}]. \quad \dots\dots (3.3)$$

Hence

$$\begin{aligned} I(x, y) &= |D(x, y)|^2 = 1 - 4 \sin^2 \frac{\alpha}{2} I_c(Y) [1 - I_c(Y)] \\ &\quad - 2 \sin \alpha \left[I_c(Y) E^*(x, y) - \frac{1}{\pi} \int_{-Y}^Y E^*(t, y) \frac{\sin c(t - x)}{t - x} dt \right] \end{aligned} \quad \dots\dots (3.4)$$

$$= I_0(x, y) - 2 I_c(Y) \sin \alpha E^*(x, y) + \frac{2}{\pi} \sin \alpha \int_{-Y}^Y E^*(t, y) \frac{\sin c(t - x)}{t - x} dt, \quad \dots\dots (3.5)$$

$$\text{where} \quad I_0(x, y) = 1 - 4 \sin^2 \frac{\alpha}{2} I_c(Y) [1 - I_c(Y)] \quad \dots\dots (3.6)$$

is the intensity distribution on a true mirror.

* Thus a mirror whose errors of figure nowhere exceed $1/20$ fringe may be regarded as nearly true in the above sense.

For $|x| > Y$, $|y| \leq H$, that is to say in the halo, we have in place of (3.2) and (3.5)

$$\begin{aligned} D(x, y) &= -[(1 - \cos \alpha) + i \sin \alpha] \frac{1}{\pi} \int_{-Y}^Y (1 + iE^*(t, y)) \frac{\sin c(t-x)}{t-x} dt, \\ I(x, y) &= |D(x, y)|^2 = [(1 - \cos \alpha)^2 + \sin^2 \alpha] \\ &\quad \times \left[(I_0(Y))^2 + \left(\frac{1}{\pi} \int_{-Y}^Y E^*(t, y) \frac{\sin c(t-x)}{t-x} dt \right)^2 \right] \\ &= 4 \sin^2 \frac{\alpha}{2} (I_0(Y))^2 \\ &= I_0(x, y). \end{aligned} \quad \dots\dots (3.7)$$

Thus the halo is unaffected by small errors of figure on the mirror.

A small local error on an otherwise true mirror corresponds to the case where $E^*(x, y)$ vanishes everywhere outside a small x interval J of length $\eta \ll H$. For each fixed y , the integral on the right of (3.5) is then $O(\eta)$ times the average value of $E^*(x, y)$ in the interval J and the intensity distribution is therefore represented to a good approximation by the equation

$$I(x, y) = I_0(x, y) - 2I_0(Y) \sin \alpha \cdot E^*(x, y) \quad \dots\dots (3.8)$$

$$= 1 - 4 \sin^2 \frac{\alpha}{2} I_0(Y) [1 - I_0(Y)] - 2I_0(Y) \sin \alpha \cdot E^*(x, y). \quad \dots\dots (3.9)$$

Equation (3.9) shows that on a small local bump the intensity is reduced by an amount very nearly proportional to $\sin \alpha$ times the height of the bump. In the neighbourhood of such a bump the 'background intensity' $I_0(x, y)$ is nearly constant, by (3.6) and (3.3). It follows that here, as in the Zernike disc test, the observed intensity change is, within the limits of visual estimation, proportional to the height changes over the bump, and the dark patch which (with a phase-retarding strip) represents the error corresponds to it in shape and intensity. What is even more important from the practical point of view, its boundary agrees exactly with the boundary of the bump, so that the mirror-maker does not find (as he does with the Foucault test) that his polisher has inexplicably trespassed outside the limits of the bump.* As a result, local figuring operations can be carried out with ease and certainty under the phase-strip test which would be impracticable under the Foucault test.

In the case where shallow irregular error covers a large part of the mirror surface, the approximation (3.9) is no longer applicable, and the more accurate formula (3.4) must be used. But since the value of the integral

$$\rho(x, y) = \frac{1}{\pi} \int_{-Y}^Y E^*(t, y) \frac{\sin c(t-x)}{t-x} dt \quad \dots\dots (3.10)$$

changes only slowly as x and y vary, except near the points $(0, \pm H)$, the intensity on and near a rapid local error is still given by an equation

$$I(x, y) = -2I_0(Y) \sin \alpha E^*(x, y) + \text{background intensity}, \quad \dots\dots (3.11)$$

and the above conclusions remain valid, except near the top and bottom of the mirror disc.

* See Ponomarev, *Bull. de l'Inst. Astronomique*, No. 30, Leningrad 1931, and Linfoot (1948).

Conclusions about the sensitiveness of the test can be drawn from (3.5) and (3.7). By (3.7), shallow errors do not appreciably affect the halo; it follows that they do not appreciably alter the total amount of light seen under test on the mirror disc, but merely redistribute it. That is to say

$$\iint_{x^2+y^2 \leq H^2} I(x, y) dx dy = \iint_{x^2+y^2 \leq H^2} I_0(x, y) dx dy. \quad \dots (3.12)$$

In the case of a small local bump, the second term on the right of (3.5) describes quantitatively the manner in which light is displaced out of the small area covered by the bump and the third term describes the manner in which the displaced light is redistributed over a horizontal band of the mirror surface. Because it is now smoothly spread over a much greater area than before, the intensity changes due to this displaced light are small in comparison with those on the bump itself.

In the case of a 'cobbled surface', a large number of irregular bumps and hollows are distributed at random over the whole region $x^2 + y^2 \leq H^2$. Since the right hand side of (3.5) remains unchanged in value when a constant is added to $E^*(x, y)$, we may suppose that the mean of $I_0(Y)E^*(x, y)$ over the surface is zero. Then it follows from (3.5) and (3.12) that

$$\iint_{x^2+y^2 \leq H^2} \rho(x, y) dx dy = 0. \quad \dots (3.13)$$

Since $\rho(x, y)$ varies slowly in comparison with $I_0(Y)E^*(x, y)$, it follows that, except near the points where $E^*(x, y)$ is abnormally small, and near the points $(0, \pm H)$, the third term on the right of (3.5) is small in comparison with the second term. That is, the resultant effect of the redistributed light is small compared with the intensity variations arising from the second term on the right of (3.5). The appearance of the cobbles under the phase-strip test is therefore, to a good approximation, a faithful representation of their true sizes and shapes, except near the top and bottom of the mirror surface.

It follows from (3.4) that the intensity changes by which the errors on a nearly true mirror show themselves under the test are proportional to $\sin \alpha$. The changes are therefore greatest when the phase-strip advances or retards the phase by a quarter-cycle. It cannot, however, be inferred from this that the test is most sensitive when $\alpha = \pm \frac{1}{2}\pi$, since the background intensity $I_0(x, y)$ also depends on α .

If the test is to be sensitive in detecting slow as well as rapid errors, α and c must be chosen so that the intensities seen under the test do not vary too much over the surface of a true mirror. In particular, $I_0(x, y)$ must not vary too much along the main diameter $y=0$, $-H \leq x \leq H$. Figure 2(a) shows the intensities seen along the main diameter and in the halo when a true mirror is tested with $\alpha = \frac{1}{2}\pi$ and with selected values of Hc , the phase-strip being set at focus. In practice, the phase-strip would be set by trial so that the disc appears as nearly as possible uniformly bright. Figure 2(b) shows the intensities along the main diameter at the focal settings which would make them as nearly uniform as possible. They were calculated by setting $E^*(x, 0) = \epsilon x^2$ in (3.5) and giving ϵ the values, found by trial, -0.065 , -0.13 , -0.21 , -0.41 , -0.70 in the respective cases $Hc = 0.815$, 1.2 , 1.4 , 1.6 , 2.0 , 2.5 . Here c is measured in u, v units, as before, so that $Hc = 3.83$ in the case of a strip just wide enough to cover the Airy disc completely. In the case $He = 25$, the above procedure gives a rather rough approximation to the

true values of $I(x, 0)$. However, it is already a rather rough approximation to use the focal settings which minimize intensity variations along the main diameter instead of over the whole disc.

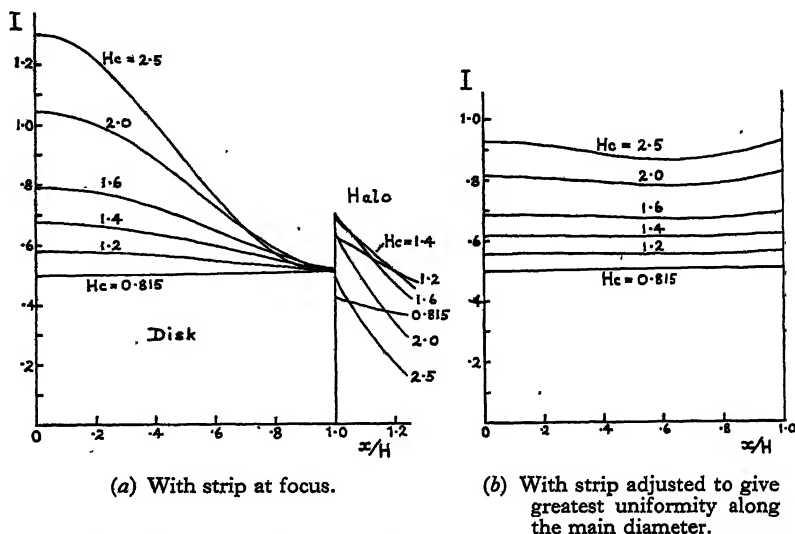


Figure 2. Intensities seen along the main diameter of a true mirror tested with quarter-wave phase-retarding strips of different widths.

From Figure 2(b) it appears that, in the case $\alpha = \pm \frac{1}{2}\pi$, Hc should not exceed about 2.0 if a true mirror is to appear uniformly illuminated along the main diameter. The width of the phase-strip is then 0.52 times the diameter of the Airy disc. Greater uniformity along the main diameter, with only a moderate decrease in intensity there, is obtained by choosing $Hc = 1.2$. The appearance of the disc with the phase-strip set at focus is then as shown in Figure 3. There is a very

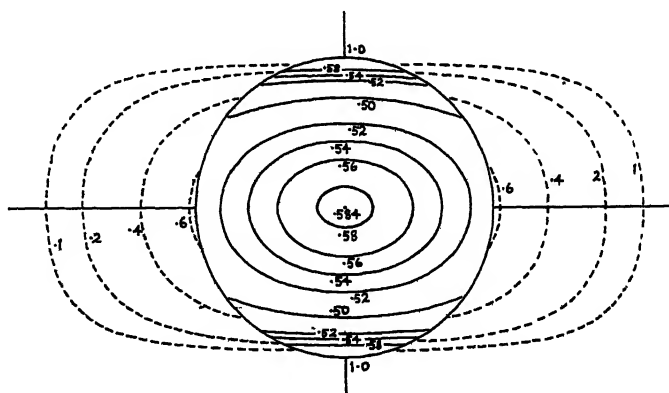


Figure 3. Intensities seen on true mirror under test with quarter-wave phase-retarding strip ($Hc = 1.2$, $\alpha = \frac{1}{2}\pi$) set at focus. The dotted isophotes show the intensities in the halo.

noticeable increase in brightness near the top and bottom of the disc, but its appearance compares favourably in uniformity with that under the Foucault test.

A strip with $\alpha = \frac{1}{2}\pi$ and $Hc = 2.0$, used at the appropriate focal setting, would not give appreciably worse uniformity over the disc as a whole than one with $Hc = 1.2$ and would be about equally useful in practice. However, in the critical

examination of weak zonal aberration something is gained by using the narrower strip, together with a slit mask exposing only a horizontal band of the mirror surface around its main diameter. By §2, the presence of the mask does not alter the intensities on the exposed portion of the mirror surface.

Increasing $|\alpha|$ to a value between $\frac{1}{2}\pi$ and π would, by (3.6), decrease the values of $I_0(x, y)$ and accentuate their fractional variation. Some improvement in the sensitivity of the test for rapid error could be obtained in this way, but at the cost of decreasing its reliability as a means of controlling slow errors.

We therefore select, as a typical special case on which to base the quantitative estimates of sensitivity in the next Section, a strip for which $\alpha = \frac{1}{2}\pi$ and $Hc = 1.2$; that is to say a quarter-wave phase-retarding strip of width approximately one third of the diameter of the Airy disc. With such a strip, giving 'negative phase contrast', a local bump appears darker than the neighbouring parts of the surface in accordance with (3.8).

§4. SLOW ERRORS; LOCAL ZONAL ERROR

A slow error whose behaviour under the test is of special practical interest is primary astigmatism, obtained on setting

$$\begin{aligned} E(x, y) &= \exp \{i(ax^2 + 2hxy + by^2)\} & (x^2 + y^2 \leq H^2) \\ &= 0 & (x^2 + y^2 > H^2). \end{aligned} \quad \dots\dots (4.1)$$

In testing a slightly astigmatic mirror, the phase-strip is adjusted so that the halo appears symmetrical* and the mirror as evenly and symmetrically illuminated as possible. It results that the strip is set centrally, but that its focal setting depends on the orientation of the mirror and on the amount of the astigmatism.

The effect of moving the phase-strip through a distance δs from O, away from the mirror, can be found by replacing $E(x, y)$ on the right of (2.5) by

$$\exp \{i\eta(x^2 + y^2)\} \cdot E(x, y)$$

and $D(x', y')$ on the left by $\exp \{i\eta(x'^2 + y'^2)\} \cdot D(x', y')$, where $\eta = \pi(\delta s/\lambda s^2)$. The resulting equation gives the new value of $D(x', y')$; on substituting for $E(x, y)$ from (4.1) and simplifying, it becomes

$$\begin{aligned} D(x', y') &= \exp \{i(ax'^2 + 2hx'y' + by'^2)\} \\ &\times \left[\frac{1}{0} - (1 - e^{-i\alpha}) \int_{x'-Y}^{x'+Y} \exp \{i(a+\eta)(2x't + t^2) + 2ihty'\} \cdot \frac{\sin ct}{t} dt \right], \end{aligned} \quad \dots\dots (4.2)$$

where the upper or lower alternative is taken according as $x^2 + y^2 \leq H^2$.

At the focal setting $\eta = -a$, (4.2) gives

$$I(-x', -y') = I(x', y'); \quad \dots\dots (4.3)$$

at other focal settings this identity does not hold. Thus there is a unique focal setting at which the astigmatic mirror appears skew-symmetrical under the test. This setting is the one which is used in practice; it will be called the preferred setting. (4.2) gives, at the preferred setting

$$I(x', y') = |D(x', y')|^2 = \left| \frac{1}{0} - (1 - e^{-i\alpha}) \frac{1}{\pi} \int_{x'-Y}^{x'+Y} \exp 2ihty' \frac{\sin ct}{t} dt \right|^2. \quad \dots\dots (4.4)$$

The integral on the right is easily evaluated in terms of the functions Si, Ci.

* Compare (3.7) above. If the amount of astigmatism is not small, the halo is skew-symmetrical by (4.3).

Of the coefficients a , h , b in (4.1), only h occurs in the expression (4.4) for the intensity at the preferred setting. It follows that a given astigmatism shows itself most strongly when the mirror is oriented with its astigmatic principal axes at angles of 45° with the direction of the phase-strip. If one of these axes is parallel to the direction of the phase-strip, the appearance of the mirror under the test is the same as that of a true mirror and the astigmatism escapes detection.*

Figure 4 shows the appearance, calculated from (4.4), of $\frac{1}{10}$ fringe of astigmatism, tested with a quarter-wave phase-retarding strip of width approximately one third of the diameter of the Airy disc, the mirror being oriented so that its astigmatic principal axes make angles of 45° with the vertical strip. In this case the

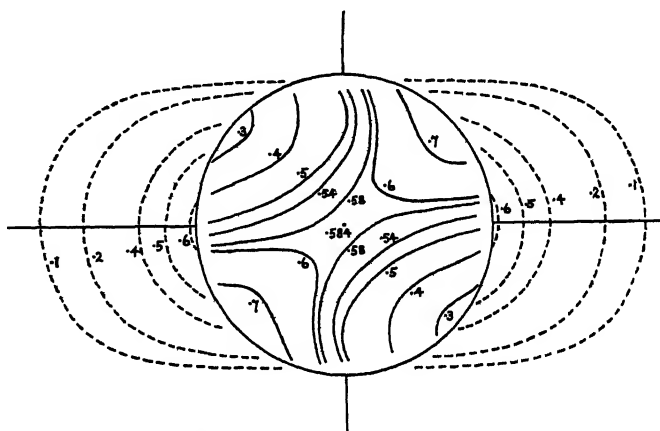


Figure 4. One tenth of a fringe of primary astigmatism ($E(x, y) = \exp \{2\pi ixy/10\}$) under test with a quarter-wave phase retarding strip ($H/c = 1.2$, $\alpha = \frac{1}{4}\pi$). The dotted isophotes show the intensities in the halo.

astigmatism can easily be recognized under the test, and it may be estimated that one-twentieth of a fringe of astigmatism should be detectable.

Of equal practical importance is the question whether a small amount of primary spherical aberration can be seen and identified without difficulty. Figure 5 shows the appearance of one tenth of a fringe of primary spherical aberration, of type

$$E(x, y) = \exp \{8\pi i(r^2 - r^4)\}, \quad r^2 = x^2 + y^2, \quad \dots\dots(4.5)$$

tested with the same strip.† From the figure it appears that a broad dark band, of mean radius approximately $0.7H$ and not much differing from a radially symmetrical ring in appearance, indicates correctly the position of the 'crest of the doughnut', while the brightening of the disc towards the centre and edge corresponds sufficiently closely with the variations in the error function $r^2 - r^4$. Thus one tenth of a fringe of spherical over-correction presents a striking and easily recognized picture under the test, and it seems safe to say that one twentieth of a fringe could be identified without difficulty.

* Compare the properties of astigmatism under the Foucault test, *Mon. Not. R. Astr. Soc.*, 1945, 105, 193.

† Calculated from (2.5). My thanks are due to the Cambridge Mathematical Laboratory for their help in computing this figure.

Finally, in order to form an idea of the sensitiveness of the test in detecting local zonal error, we consider an 'artificial' zone, one twentieth of a fringe high all over its surface, occupying the annulus

$$0.6H \leq r \leq 0.7H \quad \dots\dots\dots (4.6)$$

of the mirror surface. Evidently $E(x, y)$ has step-discontinuities on its inner and outer boundaries of the zone, as well as at the edge of the mirror, and the immediate neighbourhood of these points, as well as of the rim of the mirror, must therefore be excluded from the discussion. At the remaining points, (2.5) may be applied with

$$\begin{aligned} E(x, y) &= 1 & 0 \leq r \leq 0.6H \\ &= e^{2\pi i r/20} & 0.6H < r < 0.7H \\ &= 1 & 0.7H \leq r \leq H \\ &= 0 & r > H, \end{aligned} \quad \dots\dots (4.7)$$

where, as before, $r^2 = x^2 + y^2$. The resulting intensity distribution is represented in Figure 6, which shows that the zone will appear as a clearly marked dark ring of about half the brightness of the nearby portions of the mirror surface.

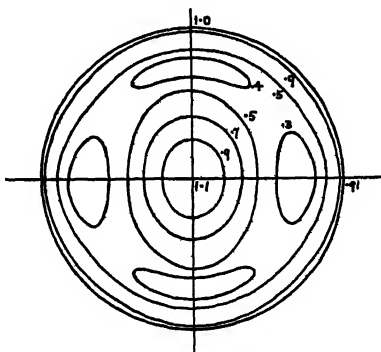


Figure 5. One-tenth of a fringe of spherical over-correction $\left(E(x, y) = \exp \left\{ \frac{2\pi i}{10} (r^2 - r^4) \right\} \right)$ under test with a quarter-wave phase-retarding strip ($Hc = 1.2$, $\alpha = \frac{1}{2}\pi$). The numbers show the relative intensities along the isophotes.

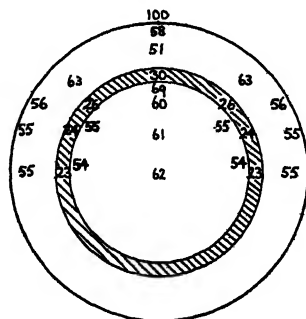


Figure 6. Flat-topped local zone
 $0.6H \leq r \leq 0.7H$,
 one twentieth of a fringe high, under test
 with a quarter-wave phase-retarding strip
 $(Hc = 1.2, \alpha = \frac{1}{2}\pi)$ set at focus. The num-
 bers show the relative intensities seen in
 their immediate neighbourhoods.

From the results of the last two sections we may infer that when carried out with a slit-source of negligible width the phase-strip test can safely be used to figure an optical surface to an accuracy of one twentieth of a fringe. Below this limit, the systematic errors of the test begin to reduce its usefulness for the control of slow errors. Rapid local errors are shown with sufficient fidelity right up to the limit of sensitiveness of the test, which for such errors is of the order of one fiftieth of a fringe.

REFERENCES

- BURCH, C. R., 1934, *Mon. Not. R. Astr. Soc.*, **94**, 384.
 LINFOOT, E. H., 1946a, *Proc. Phys. Soc.*, **58**, 759; 1946b, *Proc. Roy. Soc. A*, **186**, 72;
 1948, *Ibid.*, **193**, 248.
 ZERNIKE, F., 1934a, *Physica*, **1**, 689; 1934b, *Mon. Not. R. Astr. Soc.*, **94**, 377.

LETTERS TO THE EDITOR

On the Frequency Response of PbS Transistors

As described in a previous communication (Gebbie, Banbury and Hogarth 1950), some specimens of (deficit type) lead sulphide have been used successfully as crystal triodes, and voltage gains up to 60 have now been observed. It was of interest to compare the frequency response of such triodes with that of germanium transistors and to examine whether this response could be influenced by transverse magnetic fields as observed for n-type germanium (Bradner Brown 1949). Figure 1 shows response curves for one Ge and three PbS triodes,

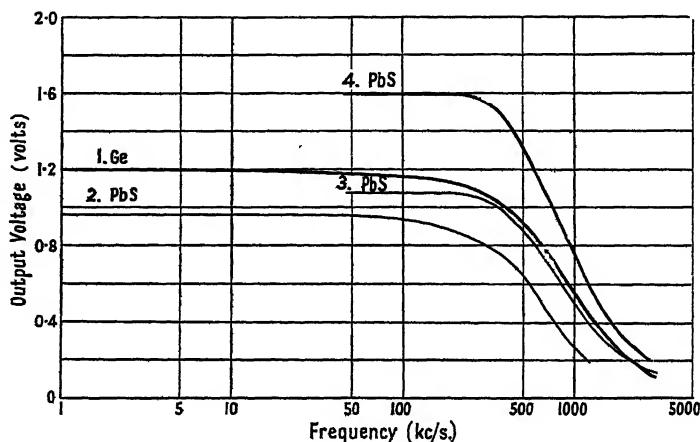


Figure 1. Frequency response of Ge and PbS transistors.

for 4 v. (1, 2, 3) and 6 v. (4) on the collector. It has been found that this frequency response of PbS triodes is dependent on the collector voltage. Thus, on a typical specimen for collector voltages of 1, 3, 6 and 8 volts, the corresponding frequencies at which voltage gain is reduced by 50% are 420, 460, 550 and 610 kc/s. Qualitatively this is expected from considerations of transit time of the charge carriers which pass from the emitter to the collector. Higher collector voltages cannot, at present, be used on PbS, but future developments in this direction would be expected to lead to higher frequency limits.

A magnetic field of about $\pm 3,700$ gauss was applied in the manner shown in Figure 2,

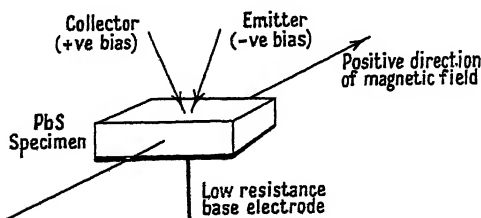


Figure 2. Direction of magnetic field relative to transistor electrodes.

and the corresponding values of voltage gain re-determined at various frequencies. The results (for 6 v. on the collector) are given in the following Table, together with similar measurements on Ge for purposes of comparison. It should be noted that these results are not directly comparable with those reported by Bradner Brown (1949), whose observations concerned *current gain*.

Voltage Output for Fields of +3,700, 0, -3,700 gauss.

Frequency (kc/s.)	PbS (p-type) : 0.05 v. input			Ge (n-type) : 0.1 v. input		
	+3700	0	-3700	+3700	0	-3700
100	1.22	1.19	1.16	0.97	1.02	1.07
200	0.94 ₀	0.92 ₀	0.90 ₀	0.96	1.01	1.07
300	0.68 _s	0.67 ₀	0.65 _s	0.87	0.93	0.97
500	0.37 ₀	0.36 ₀	0.35 ₀	0.68	0.73	0.76
700	0.25 ₀	0.24 ₀	0.23 ₀	0.52	0.57	0.59
1000	0.16 _s	0.16 _s	0.16 ₀	0.38	0.40	0.42
1500	0.12 _s	0.12 _s	0.12 ₀	0.24	0.26	0.28
2000	—	—	—	0.17	0.18	0.20

According to the familiar transistor concepts, the magnetic field is expected to affect the average path length of the charge carriers. This length should be reduced by fields which tend to deflect the current flow from emitter to collector towards the surface. This would reduce the transit time which, together with the life-time of the carriers, limits the gain and its frequency response.

Although the effect of the magnetic field is small, it can be seen from its polarity that the charge carriers injected by the emitter into the PbS crystal must be electrons. The mechanism of this triode action is expected to be closely analogous to that of the Ge p-type transistor as reported by Pfann and Scaff (1949).

Thanks are due to Professor R. W. Ditchburn for placing research facilities at our disposal and to Dr. C. A. Hogarth for much helpful advice.

Department of Physics,
The University,
Reading.
20th April 1950.

P. C. BANBURY.
H. K. HENISCH.

BRADNER BROWN, C., 1949, *Phys. Rev.*, **76**, 1736.

GEBBIE, H. A., BANBURY, P. C., and HOGARTH, C. A., 1950, *Proc. Phys. Soc. B*, **63**, 371.

PFANN, W. G., and SCAFF, J. H., 1949, *Phys. Rev.*, **76**, 458 (A).

REVIEWS OF BOOKS

Medical X-Ray Protection up to Two Million Volts, NATIONAL BUREAU OF STANDARDS. Handbook 41. Pp. 41. (Washington: Superintendent of Documents, U.S. Department of Commerce, March 1949.) 15 c.

The present handbook supersedes a previous one of the National Bureau of Standards, last revised in 1936. It contains general sections, rules for specific applications in various kilovoltage regions and particularly details of electrical protection. The protective measures are based upon a permissible dosage-rate of 0.3r/week in place of 0.5r/week adopted in the recently issued Recommendations of the British X-Ray and Radium Protection Committee (October 1948), but it appears that the two conditions of measurement are different in the two cases, and over a wide range of radiations the numbers are essentially the same. While the recommendations are intended primarily for the protection of the radiation worker and not for the patient, the preface very rightly emphasizes the importance of caution and restraint in the use of x-rays for diagnosis and the treatment of non-malignant conditions. The genetic effects of radiation increase in importance as a larger fraction of the population is exposed to ionizing radiation, and diagnostic x-ray examinations become more and more frequent. Moreover, the employment of a large number of people in industries making use of ionizing radiations calls for, and is being given, serious thought. The present handbook merely hints at these problems and describes the consultative and advisory structure set up in the United States to deal with them.

Study of the recommendations shows that they are thorough and very detailed. A particularly useful appendix gives radiation attenuation curves in lead and concrete for beams of x-rays produced at voltages up to 2,000 kv. with worked examples illustrating practical calculations of protective requirements.

The handbook should be in use in all medical institutions using high voltage x-rays. It also contains much information which is appropriate to industrial organizations making use of these radiations.

W. V. M.

CONTENTS FOR SECTION A

	PAGE
Mr. SURAJ N. GUPTA. Theory of Longitudinal Photons in Quantum Electrodynamics	681
Prof. M. H. L. PRYCE. On the Energy of the Alpha-Particles from Radioactive Nuclei	692
Dr. W. E. MOFFITT. Excited Electronic Levels in Conjugated Molecules—IV: Symmetrical Cyanine Dyes	700
Mr. J. EWLES and Dr. C. CURRY. Resolution and Analysis of Low Temperature Luminescent Spectra of Bi and Pb Activated Solids of Simple Crystal Structure	708
Dr. A. A. SABRY. Influence of Dipole-Dipole Coupling on the Dielectric Constant of Dipolar Substances	716
Dr. P. WRIGHT. The Effect of Occluded Hydrogen on the Electrical Resistance of Palladium	727
Dr. R. S. TEBBLE, Dr. I. C. SKIDMORE and Dr. W. D. CORNER. The Barkhausen Effect	739
Dr. K. H. STEWART. Domain Wall Movement in a Single Crystal	761
Prof. A. RUBINOWICZ. Sommerfeld's Polynomial Method Simplified	766
Dr. M. W. FEAST. Investigation of the Spectrum of the High Voltage Arc in Carbon Dioxide: the CO Flame Spectrum	772
Letters to the Editor:	
Dr. F. D. S. BUTEMENT. Radioactive ^{185}Er	775
Mr. E. R. RAE, Mr. J. G. RUTHERGLEN and Mr. R. D. SMITH. Proton Capture Radiation of Fluorine-19	775
Prof. H. FRÖHLICH. Isotope Effect in Superconductivity	778
Dr. A. G. GAYDON and Dr. H. G. WOLFHARD. Excitation of Spectra in the Inner Cones of Flames	778
Corrigenda	780
Reviews of Books	780
Contents for Section B	783
Abstracts for Section B	783

ABSTRACTS FOR SECTION A

Theory of Longitudinal Photons in Quantum Electrodynamics, by SURAJ N. GUPTA.

ABSTRACT. The radiation field is quantized by introducing four types of photons—two transverse, one longitudinal, and one scalar. The scalar photons are treated by using an indefinite metric, and it is found necessary to modify the usual supplementary condition slightly. The present theory offers a justification for the symmetrical treatment of the four components of the electromagnetic potential, recently applied by a number of authors, and proves to be very convenient in applications. The results of physical interest, however, are the same as obtained from the ordinary formulation.

On the Energy of the Alpha-Particles from Radioactive Nuclei, by M. H. L. PRYCE.

ABSTRACT. When the energy of α -particles emitted by radioactive nuclei is compared with the energy calculated from the usual semi-empirical formula for nuclear binding energies, systematic variations ranging from 4.5 Mev. to -1.0 Mev. are found, which exhibit in a very clear way the special importance of configurations containing 82 protons and/or 126 neutrons. The general features are readily explained in terms of the shell model. The analysis can be used to predict unknown α -energies. Calculated α -energies are tabulated for the heavy nuclei ($Z \geq 78$) and for nuclei in the rare-earth region.

Excited Electronic Levels in Conjugated Molecules—IV: Symmetrical Cyanine Dyes, by W. E. MOFFITT.

ABSTRACT. Topics in the electronic theory of the colour of symmetrical cyanine dyes have been discussed by means of resonance formalism. Our conclusions for the formamidinium ion have been confirmed by use of the molecular orbital method. As a result of this analysis, certain modifications in the qualitative symbolism of colour theory have been put forward in the hope that the factors governing absorption may be more clearly understood.

Resolution and Analysis of Low Temperature Luminescent Spectra of Bi and Pb Activated Solids of Simple Crystal Structure, by J. EWLES and C. CURRY.

ABSTRACT. Experimental arrangements are described by which clearly resolved spectra have been obtained of the luminescence, at the temperature of liquid air, of seven different phosphors of simple crystal structure (cubic) and of mixed crystals. The main peaks in each case may be arranged in a band series. The vibration frequency in the lower state is nearly the same for the same host lattice with different activators. The energy of the electronic transition is determined both by the activator and the host crystal and there is a progression of decreasing energy in the series of simple crystals MgO, CaO, SrO, CaS, and in the order CaO, CaO-SrO with mixed crystals. The most probable transitions are from the lowest level of the upper state. There is evidence that self-excitation occurs and for the view that the luminescent centres are large.

Influence of Dipole-Dipole Coupling on the Dielectric Constant of Dipolar Substances, by A. A. SABRY.

ABSTRACT. Kirkwood's formula for dielectric constant, ϵ , gives it in terms of a correlation coefficient which depends on the probability that any two dipolar molecules have a certain configuration.

The main object of this paper is to calculate the configurational probability of the dipole orientations; this is dealt with in the first part of the paper by considering two dipoles and replacing the rest by a homogeneous dielectric. The dipolar molecules are regarded as point dipoles and the effect of other forces on the orientation of the molecules is neglected.

In the second part of the paper the calculation of the correlation coefficient appearing in Kirkwood's dielectric constant formula is carried out. This correlation coefficient is evaluated for water, and the value of ϵ then obtained from Kirkwood's formula is in fair agreement with the observed value.

The Effect of Occluded Hydrogen on the Electrical Resistance of Palladium, by P. WRIGHT.

ABSTRACT. Agreement between the results of theoretical and experimental investigations of the absorption of hydrogen by palladium indicates that, at least at higher temperatures, the hydrogen is contained in solid solution. Apparatus for observing the changes in electrical resistance of palladium during absorption and evolution of hydrogen is described, with which resistance-pressure isotherms from 75° to 150° C. were obtained for pressures increasing and decreasing between zero and atmospheric. Factors affecting the resistance of the Pd-H system are discussed and an explanation of the form of the isotherms is given; a decrease in resistance which is observed, when hydrogen is added to the alloy in the β phase, is attributed to completion of the d shells of palladium.

The Barkhausen Effect, by R. S. TEBBLE, I. C. SKIDMORE and W. D. CORNER.

ABSTRACT. An account is given of further experiments on the Barkhausen effect in cylindrical ferromagnetic specimens. The form of the decay of induction following a Barkhausen discontinuity has been investigated and there is reasonable agreement between the theory given and the experimental results. As a result of this investigation it has been

possible to satisfy the different conditions which are necessary for the detection of the discontinuities so as to obtain (a) accurate reproduction of the form of the discontinuities and (b) maximum sensitivity in the counting experiments. The number and size of the discontinuities are measured, and an estimate is made of the contribution of the discontinuous processes to the total change in magnetization over the relevant part of the hysteresis curve. Discontinuous changes in magnetic moment greater than 0.3×10^{-6} E.M.U. (corresponding to a minimum volume of 0.9×10^{-10} cm³ for iron) account for from 47% to 86% of the total change in magnetization. This is considered in relation to associated work on reversible changes, and the nature of a Barkhausen discontinuity is briefly discussed with reference to other investigations on domain structure.

Domain Wall Movement in a Single Crystal, by K. H. STEWART.

ABSTRACT. A single crystal of silicon-iron was shaped so as to encourage the formation of large ferromagnetic domains. Its magnetization curve was found to contain a large Barkhausen jump, indicating reversal of magnetization in a volume of the order of $1/10$ cm³. The rate of change of magnetization was found to be proportional to the excess of the applied magnetic field over a certain critical field, and could be made very slow. It is suggested that the rate of change is controlled by the braking effect of eddy currents on the movement of domain walls.

Sommerfeld's Polynomial Method Simplified, by A. RUBINOWICZ.

ABSTRACT. From two earlier papers by the author it follows that eigenfunctions obtainable by Sommerfeld's polynomial method contain either a Riemann P -function or a degenerate form of it when the fundamental interval is either finite or infinite respectively. Comparing the differential equation for the P -function or its degenerate form with the differential equation which we get for P from the original differential equation of the eigenvalue problem, we obtain standard formulae for the constants determining the eigenvalues and eigenfunctions. The calculations necessary to give the complete solution of an eigenvalue problem which can be treated by Sommerfeld's polynomial method, are thus reduced to a minimum of differentiations and simple algebraic operations. It is also possible, using appropriately chosen P -functions, to obtain the eigenfunctions directly in the required form. The method yields also eigenfunctions for eigenvalues in the continuous spectrum.

Investigation of the Spectrum of the High Voltage Arc in Carbon Dioxide: the CO Flame Spectrum, by M. W. FEAST.

ABSTRACT. The spectrum emitted by a high voltage arc in CO₂ at atmospheric pressure has been investigated in the region 2000 Å. to 9000 Å. It was found to consist, mainly, of a continuum and the 'CO flame bands'. A study of the spectrum at different dispersions and on photographic plates of various kinds shows that although the Schumann-Runge O₂ bands do form a part of the spectrum they are relatively weak compared with the bands attributed to CO₂. The excitation of the Schumann-Runge O₂ system in flames is discussed.

THE PROCEEDINGS OF THE PHYSICAL SOCIETY

Section B

VOL. 63, PART 8

1 August 1950

No. 368 B

New Localized Multiple-Beam Interference Fringes formed with Curved Thin Sheets

BY S. TOLANSKY AND N. BARAKAT

Royal Holloway College, University of London

MS. received 25th October 1949, and in amended form 7th February 1950

ABSTRACT. A description is given of the mode of formation and optical properties of some new sharp localized multiple-beam fringes produced by cylindrically curved thin sheets which are silvered upon both sides. Illumination with strictly parallel monochromatic light leads to sharp line fringes lying in a plane passing through the centre of curvature and perpendicular to the direction of incidence. An approximate theory is developed and checked by observation. The use of mica as the film leads to interesting birefringent doubling effects. These are complicated by differential phase change effects occurring on reflection at the silver surfaces. Fringe systems have been examined both in reflection and in transmission and related systems obtained with white light used together with a spectrograph. Cleavage step phenomena are revealed by the mica with high precision and high dispersion. It is proposed to name the monochromatic system 'fringes of equal tangential inclination'.

§ 1. ORIGIN OF THE FRINGES

IN this paper a description is given of the mode of formation and general properties of some new localized multiple-beam interference fringes which are formed by thin sheets of parallel-sided transparent material silvered on both sides and then bent into cylindrical curvature.

A suitable material is a thin cleavage sheet of muscovite mica, and since the use of such material leads in addition to some interesting complex birefringent effects, most of the experiments reported here have been made with varieties of mica.

The fringes in their simplest form are produced in transmission with monochromatic light, but variants, such as reflection fringes and white light modifications both in transmission and in reflection, have also been studied. The selected thin sheet of material is coated on both sides with high reflectivity silverings of the type generally adopted for multiple-beam interferometry (Tolansky 1948 a) and viewed with the simple optical arrangement shown in Figure 1. S is the source, C a condenser and F a filter.

A parallel beam of monochromatic light is incident on the convex face of the mica. Related but different effects result if mica is turned so that the light falls on the concave face. A suitable thickness, t , of mica is some 1/50 mm. and a suitable cylindrical curvature has a radius, R , of perhaps 2 cm. Fringes can also be obtained over relatively wide ranges of both t and R . An example taken with the green mercury line is shown in Plate I (*a*).

The origin of formation of the fringes is shown by Figure 2, where for simplicity refraction effects within the interference film have been disregarded. To a close approximation both faces can be considered to have the same radius of curvature. Parallel incident light at any specific point A is incident on the film at an angle θ which is determined by the distance XY above the axis of

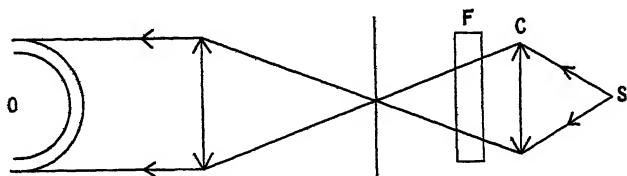


Figure 1.

symmetry XO. Providing t is small compared with R (in the case quoted the ratio is 1:1,000) then for any particular incident ray, YA, the successive multiple reflections meet at a point M and there interfere. This can be demonstrated either by ray tracing or by calculation. If θ , R and t are such that the optical path difference between successive beams (including phase change effects) is an integral number of waves $n\lambda$, then reinforcement occurs at M and a bright fringe results. Clearly at a further distance from XO there is another ray with a suitable angle of incidence leading to another fringe with path difference $(n-1)\lambda$. Hence a succession of localized fringes form and owing to the multiple-beam effect these are very sharp.

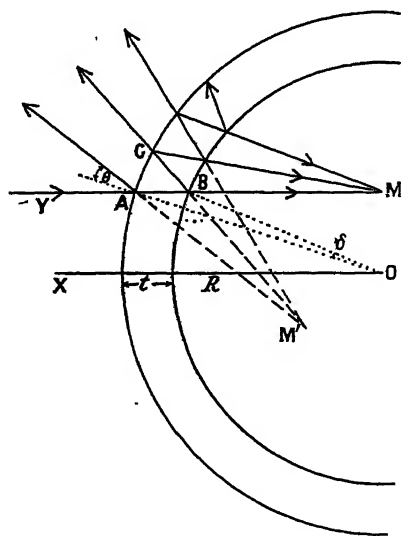


Figure 2.

In this introductory report in which it is intended to survey the subject only, a simplified geometrical theory will be developed, justifiable approximations being used, and in a later communication the theory will be treated fully and applications to birefringent films examined critically. A preliminary notice has already been communicated elsewhere (Tolansky and Barakat 1948).

§ 2. FRINGE LOCALIZATION

It is readily demonstrated by the following simple geometrical construction that to a close approximation, when t is small, successive fringes are localized on a plane passing through O, the centre of curvature, and perpendicular to the symmetry axis XO. From Figure 2 it follows that

$$BM = BC \frac{\sin 2\theta}{\sin 2\delta}; \quad BC = AB = R \frac{\sin \delta}{\sin \theta};$$

therefore

$$BM = R \frac{\cos \theta}{\cos \delta}.$$

Since the angle δ is very small, BM is approximately equal to $R \cos \theta$ and the angle OMB is very nearly $\frac{1}{2}\pi$. Thus the locus of the point M for various angles of incidence is the straight line passing through the centre of curvature O at right angles to the axis XO.

It may be pointed out that the strict condition of incident parallel light, with resulting plane localization, extreme sharpness and linear shape, makes the fringes described here completely different from the broad, non-localized fringes of shape varying with the observer's position, found with unsilvered curved mica plates by Ramàn and Rajagopalan (1939). The source we have used is a pinhole of diameter $\frac{1}{2}$ mm. at the focus of well-corrected lens of focal length 15 cm. and diameter 6 cm.

If chromatic dispersion effects be disregarded then to a close approximation the condition of interference is controlled by the relation $n\lambda = 2\mu t \cos \theta$. Strictly speaking this should be replaced by the modified refraction expression which holds for either a Fabry-Perot with a dispersing medium or for a Lummer plate. This more exact treatment leads to corrections of but a few per cent. Furthermore, a detailed analysis shows that one should write $n\lambda = 2\mu t \cos \theta + \Delta$ in which Δ is a correcting term which is primarily a function of θ and δ where δ is the small angle between the radii for successively reflected rays (see Figure 2), but this too can be shown to be a secondary matter.

Measurement shows that up to angles of 60° the fringes are critically localized in the predicted plane, but beyond this up to the limiting 90° value there are slight deviations. The following table shows a comparison between the predicted linear localization and that observed for a particular mica sample with $t = 1/50$ mm., $\mu = 1.59$ and θ in the neighbourhood of 30° . The value of R was varied and the distance between the film surface and fringe localization measured. All figures are in centimetres.

R	Localization	
	Predicted	Observed
1.5	1.30	1.32
2	1.73	1.68
3	2.60	2.60
4	3.46	3.4
5	4.33	4.2

The observed localizations agree with prediction to within a fraction of a millimetre. The larger the R value the broader is the fringe and then some small experimental uncertainty in localization arises.

§ 3. PROPERTIES OF THE FRINGES

The fringes obtained with a selected piece of high quality muscovite are shown in Plate I(a). Attention may be drawn to the following features: (i) the fringes are sharply localized in the predicted plane; (ii) they are very narrow, the outer orders becoming progressively sharper; (iii) successive orders clearly follow some simple geometrical law of separation; (iv) the fringes are double and the doubling exhibits an unexpected variation with order. These features will now be discussed in turn, the consideration of the doubling being left until later.

(i) The localization in the predicted plane is as exact as can be measured with a microscope having a small depth of focus.

(ii) A striking characteristic is the unusual sharpness of the fringes for a given silvering. The reason lies in the increased reflectivity resulting from the large values of the angles of incidence θ which vary over the whole range from 0 to 90° . It is this increase in θ which is directly responsible for the marked sharpening up of the outer orders by producing high effective Fresnel reflecting coefficients at the metallic surfaces. To appreciate the sharpness fully one must focus on the true surface of localization, not on the approximate plane. It will be shown later that the two components are plane polarized mutually perpendicularly, hence the doublets can be made to appear as single fringes with a polarizing device. Figure 3 shows a microphotometer trace across a sequence of single fringes. The

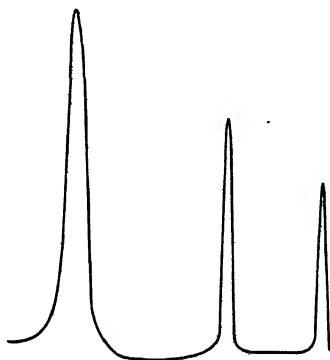


Figure 3.

outer orders are not only sharper, but fall off in intensity too. This is due to several causes, the major ones being the rapid increase in absorption by the silver films and by the mica itself with increase in θ .

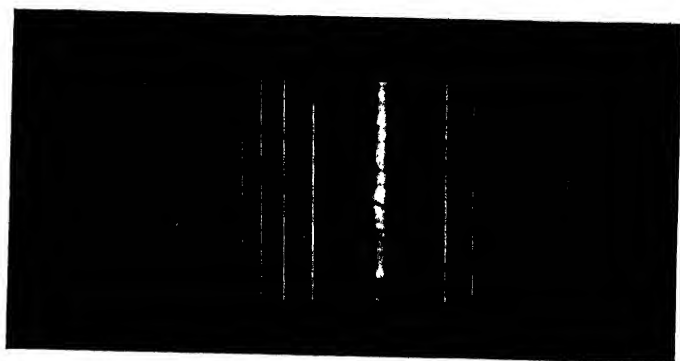
(iii) It will suffice here to indicate a simple derivation of the fringe separation law for successive orders, dispersion effects being disregarded.

The order of interference, n_0 , at the centre of the system is a maximum and is given by $n_0\lambda = 2\mu t$. Consider a symmetrical system with fringes on either side of the axis XO, then the distance between a pair of symmetrically situated fringes on either side of this axis will be called 2ρ , the 'diameter' of the fringes. If n is the order of interference for the p th fringe, counting from the centre outwards, then by analogy from the well-known corresponding Fabry-Perot theory one obtains the following: Refraction can no longer be disregarded. Let r be the angle of refraction within the film; then as $n_0 = 2\mu t/\lambda$, the order of interference at the centre, the order n for the refracting angle is $n = n_0 \cos r$; $\sin \theta = \rho/R$ so that

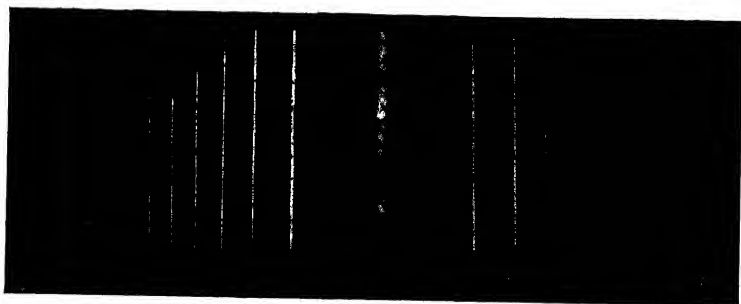
$$\frac{n}{n_0} = \left[1 - \frac{\rho^2}{u^2 R^2} \right]^{1/2}.$$



(a)



(b)



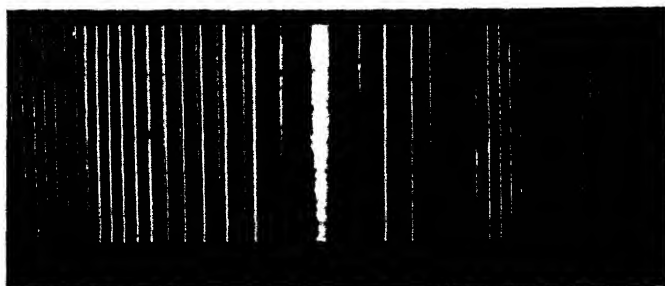
(c)



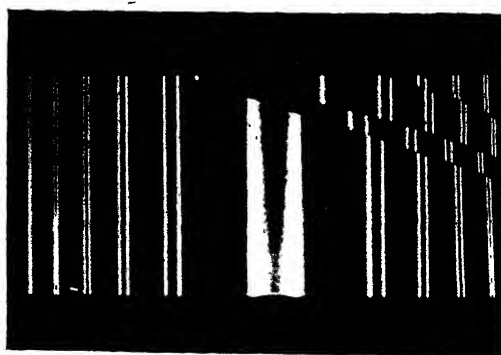
(d)

PLATE I.

To face page 548



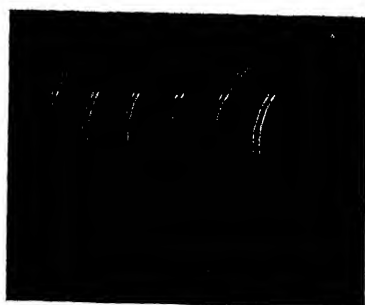
(a)



(b)



(c)



(d)

In general the order of interference n_0 at the centre will be non-integral; let it therefore be written $n_0 = n_1 + \epsilon$, where ϵ is a fraction and n_1 the order of the first fringe. The order of the p th fringe will thus be $n = n_0 - (p + \epsilon - 1)$.

Hence

$$\left(1 - \frac{p + \epsilon - 1}{n_0}\right)^2 = 1 - \frac{\rho^2}{\mu^2 R^2}.$$

As p is small compared with n_0 , then sufficiently closely

$$\frac{2}{n_0}(p + \epsilon - 1) = \frac{\rho^2}{\mu^2 R^2}$$

giving:
$$\rho = R \left[\frac{\mu \lambda}{t} (p + \epsilon - 1) \right]^{1/2}.$$

The scale of the fringe pattern is therefore directly proportional to R and inversely proportional to $t^{1/2}$.

This formula has been confirmed by measurement and shown to hold closely. Plates I(b) and I(c) illustrate the fringes given by a piece of mica acting as isotropic in which R has the respective values 1.5 and 2 cm. By assuming for simplicity that n_0 is integral (i.e. $\epsilon = 0$) the scale of the pattern can be readily calculated. Thus for example for $R = 5$ cm. and $t = 1/50$ mm. the diameter of the fifth fringe ($\lambda = 5460$ Å.) is almost exactly 4 cm.

The fringes are therefore very easy to see with the naked eye if they are allowed to fall on a ground-glass screen placed in the plane of localization. They can of course be projected and enlarged if so required for demonstration purposes.

§ 4. NOMENCLATURE

It is clear that these fringes are intimately related to the two classical types of multiple-beam interference fringes, namely fringes of *equal thickness* and fringes of *equal inclination*; it is desirable to classify them so as to retain this relationship, and several such methods of classification could reasonably be adopted.

Since one way of regarding them is the recognition of the fact that each fringe is formed as the locus of rays which on entering the system make the same angle of inclination with the tangent at the point of entry, we propose that these fringes be called 'fringes of *equal tangential inclination*'.

This terminology fits all cases and if, for example, a spherical shell replaces the cylindrical film (e.g. a blown thin curved film of glass of uniform thickness) then ring-shaped localized fringes arise, since the equal tangential loci then lie on rings.

The relationship to fringes of equal inclination is clear, for the fringes under discussion move off to infinity and become identical with fringes of equal inclination when R becomes infinite, i.e. when the sheet is *plane* parallel.

§ 5. CLEAVAGE EFFECTS

The use of mica as the material leads to an interesting application linked with the cleavage properties: it has already been established that mica sheets usually cleave in steps which are integral multiples of the crystal lattice spacing (20 Å.). If the selected mica exhibits cleavage steps the result on flexing is the production of fringes as shown in Plate I(d) in which the cleavage discontinuities are immediately evident. For let the mica thickness in a given region suffer an increment dt due to a cleavage step. Then near the centre of the system

$dt = (\lambda/2\mu) dn$, where dn is the fractional change of order produced by the step dt at a certain angle of incidence. With monochromatic light the direction of displacement and any integers of fringe overlapping cannot be easily decided, but if several distinct wavelengths are used the direction and exact number of orders can both be determined.

It is clear that the fringes offer a new precision method for the measurement of cleavage steps because a change of a whole order ($dn=1$) requires a change of thickness of $\lambda/2\mu \cos r$, i.e. about 1700 Å. near the centre (in the green). Since the fringe sharpness permits the detection of one-hundredth of an order displacement, then steps of 17 Å. can be detected.

This property does not depend on the absolute value of t except in so far as the latter affects the whole scale. This is a marked advantage over the fringes of equal chromatic order. Attention is drawn to the possibility of applying this technique to the study of small steps in thin films, such as, for example, organic films, or metal films deposited on a base such as mica.

Particularly valuable features in this connection are: (a) the high fringe dispersion available without any costly auxiliary equipment and (b) the flexibility in dispersion which is controllable by varying R . The latter can be adapted to the value of t . For example sharp fringes have been obtained with a thin film of MgF_2 less than a light wavelength thick, deposited by evaporation between two silver films.

§ 6. DIFFERENTIAL PHASE CHANGE EFFECT

It has been shown in former studies (Tolansky 1948 b) that when the localized multiple-beam fringes given by an air wedge are formed with non-normal incidence, the fringes split into doublets, the two components being plane polarized mutually perpendicularly. Both become sharper with increasing incidence but one much more rapidly than the other, and at the same time the sharper fringe weakens rapidly relative to its slightly broader companion. It has been fully established that this is due to the differential phase change which takes place when mutually perpendicular vibrations of light are reflected at non-normal incidence from a metal, and is in fact the same effect that leads to the elliptic polarization of such reflected light.

Because of the severe bending of the silvered mica sheet, considerable angles of incidence are involved and thus inevitably the outer orders must show the differential phase change doubling. This is an effect which increases at such a rate that it becomes perceptible at $\theta = 25^\circ$ and increases to about half an order for higher values of θ .

§ 7. BIREFRINGENCE DOUBLING

When using muscovite mica interesting complex effects arising from the biaxial crystal make their appearance, as for example in Plate II (a). The fringes are double, mutually perpendicularly plane polarized. This fringe doubling diminishes to a small value as the order increases and then begins to increase once more. Over the whole range the outer component becomes progressively sharper and weaker and then effectively vanishes. A further example, Plate II (b) shows a piece of mica which exhibits both the complex doubling and also marked cleavage steps. In the discussion that follows it will be seen that the observations

are completely accounted for in that they represent a combination of birefringent doubling due to the crystal, on which is superposed differential change doubling due to the silver.

In general the splitting of any incident ray by the crystal into two mutually perpendicularly polarized rays travelling with different velocities leads to two systems of interference fringes. Leaving out for the moment the phase change effect, then the separation between any pair of members belonging to the two systems, but of the same order of interference, is due to the difference in the refractive indices of the two vibrations. Thus for a biaxial crystal there will be two directions in which the difference in refractive indices is zero. These are in the neighbourhood of the positions indicated on Plate II (*a*) where the separation is reduced to a minimum. It is obvious that the actual angular location of the zero positions within the fringe system are decided by the direction of flexure of cylindrical curvature. Plate II (*a*) refers to a specimen bent along the direction of intersection of the optical axial plane with the cleavage plane.

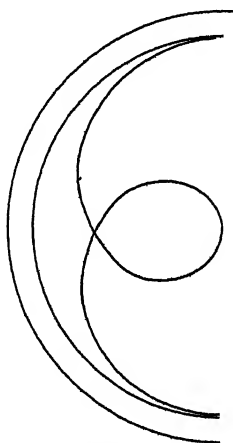


Figure 4.

Thus the angular separation between the two positions of minimum separation will give double the apparent optic angle of the mica if the differential change of phase at reflection mica-silver is allowed for. Measurements show this to be in fact the case, but will be reserved for the later more detailed discussion on birefringence.

§ 8. THE REFLECTED SYSTEM

As in all multiple-beam interference phenomena the transmitted fringes have an associated 'complementary' reflected system in which the fringes appear as fine dark lines on a bright broad background. However, the fringes under discussion are unique in that the localization of the reflected system is different in type from that of the transmitted system. Figure 2 shows that the reflected fringes which form are virtual arising from diverging beams and appear to be localized in a different position from the transmitted system. Figure 4 shows the localization derived from ray tracing. Because of the shape of the surface of localization it is difficult experimentally to confirm this localization with the exactness possible for the transmitted system.

It is of interest to note that reasonably sharp fringes can be obtained in reflection *without* silvering owing to the high Fresnel reflecting coefficients of the uncoated mica consequent on the high angles involved; Plate II (c) is an example. The fringes are relatively sharp for a 'two-beam' interference phenomenon. Actually of course there must be several effective beams to each fringe.

§ 9. FRINGES OF EQUAL CHROMATIC ORDER

In conclusion a brief indication will be given of the possibilities of investigation opened up by a study of the associated phenomena using white light, thus producing fringes analogous to the white light fringes of equal chromatic order. The system is illuminated with parallel white light and an image of the plane of localization projected on to the slit of the spectrograph. A typical result is shown in Plate II (d) in which a doubly silvered mica sheet was used. This shows the birefringence of the fringes which has several advantages over the monochromatic system. Any existing chromatic variations can be obtained over a more extensive wavelength range from a single photograph. More useful still is the ease with which order allocations can be made and the angular positions of intersections of the birefringence components established. In the case of the monochromatic fringes this may require a series of observations with different wavelengths if ambiguities are to be avoided; the white light fringes supplying inherently a continuous number of wavelengths do not suffer from this defect.

§ 10. OTHER APPLICATIONS

The experiments described here have been made with muscovite mica but there are many other distinct possibilities. Already some preliminary studies have been made on thin evaporated magnesium fluoride films. When uniform thin films of such materials are used as interference films the Fizeau fringe system is simply a field of uniform tint and this is the basis of interferometric colour filters. Thus it is difficult to obtain any numerical information from any local tint variations of such a film arising from changes in thickness, and furthermore, a determination of the thickness itself is a difficult matter unless the film growth from zero thickness has been observed. If such a thin film is built up on silvered mica, and then again silvered, flexing leads to the localized fringes even if the thickness of the film is appreciably less than half a light wavelength. With such a film we have obtained both excellent monochromatic and white light fringes, and in view of the sensitivity of fringe position to thickness, this result offers a further means of studying such films.

Another possible field of application is to the study of the organic layers which it is possible to put down on a silvered mica surface.

Since the fringes can be formed in reflection it should be possible to apply the technique to the examination of the surfaces of metallic foils either by coating these with a thick transparent film, which is then silvered, or alternatively, by pressing into contact with a silvered glass cylinder. It is of course certain that t will vary considerably in the latter case, hence we are already exploring the nature of the fringes formed by flexion of slightly wedged film instead of a parallel-sided film.

REFERENCES

- RAMAN, C. V., and RAJAGOPALAN, V. S., 1939, *J. Opt. Soc. Amer.*, **29**, 413.
TOLANSKY, S., 1948 a, *Multiple-beam Interferometry of Surfaces and Films* (Oxford: Clarendon Press), p. 26; 1948 b, *Ibid.*, p. 38.
TOLANSKY, S., and BARAKAT, N., 1948, *Nature, Lond.*, **162**, 816.

The Monocentric Schmidt-Cassegrain Cameras

By P. A. WAYMAN

The Observatory, Cambridge University

MS. received 23rd September 1949

ABSTRACT. Formulae are obtained for the leading coefficients in the plate-profile expansion of the monocentric Schmidt-Cassegrain cameras. These are used to derive expressions for the monochromatic and chromatic aberrations of the general monocentric system, and spot-diagrams are included to illustrate the image quality in a typical case. A comparison is made with an ordinary Schmidt camera, which has the same aperture and the same depth of figuring on the aspheric plate.

§1. INTRODUCTION

IN the classical Schmidt camera the focal length is never more than half the total tube-length, and in large astronomical systems this will increase the cost of mounting and dome. Among other advantages, more compactness is obtained by the use of a convex secondary mirror, as was originally suggested by Baker (1940) and Burch (1942), to form the so-called Schmidt-Cassegrain systems. Linfoot (1944) has carried out a general discussion of such systems, either aplanats or anastigmats, and with flat or curved field-surfaces, using the method of plate-diagram analysis.

In a practical design the best results may be obtained by retaining or introducing small controlled amounts of Seidel error in order to balance the higher aberrations. This balance can be made experimentally with the help of suitable optical tests or can be predicted by extensive ray-tracing. In any case a knowledge of the higher aberrations of the unbalanced design in a general analytical form is a valuable pointer towards the selection of a particular type of Schmidt system.

An expression for the leading aberrations of the ordinary Schmidt camera can be obtained explicitly in a relatively simple way by a method originally due to Carathéodory (1940).^{*} Use is made of the fact that the mirror and the field-surface are parts of concentric spheres whose centre in the case when the aspheric surface of the corrector plate is turned towards the mirror is the pole of that surface. The angular aberrations are then, to a first approximation, the same as the deviations from parallelism of a pencil of rays issuing from a point of the image surface, after reflection at the mirror and refraction at the corrector plate. These deviations can be calculated if assumptions are made of axial stigmatism in the light of a particular wavelength and of the monocentric property of mirror and field-surface about the pole of the aspheric surface.

A corresponding procedure cannot be adopted for the general Schmidt-Cassegrain system because the simplifying condition of concentric mirrors and field-surface no longer exists. The special case of the monocentric Schmidt-Cassegrain cameras can however be dealt with in an analogous way because here the two spherical mirrors have a common centre through which the aspheric surface of the corrector plate passes (see Figure 3). The field-surface, which, if it is to be fully accessible, must lie well behind the front surface of the primary mirror,

^{*} Linfoot (1949) has developed this method fully in an elegant and comprehensive treatment of the classical Schmidt camera.

is a sphere of radius equal to the focal length of the system and is concentric with the mirrors. The ratio of the radii of the two mirrors is, in practice, limited to a rather small range by consideration on the one hand of accessibility of image-surface and on the other of the obstruction which the secondary mirror causes to pencils entering the system.

In §2 the equation of the plate-profile is obtained in the form of a power series in r , the zonal radius in the aperture. This expansion is carried up to the sixth-power terms, which is far enough to give a high degree of axial stigmatism, and in §3 use is made of the leading terms of this expansion to obtain an expression for the monochromatic and chromatic aberrations of the system. In §4 spot diagrams of the image points are given for an $f/3.5$ system covering a five-degree field in light of different wavelengths. In addition, a comparison is made with an ordinary Schmidt camera which, using the same aperture diameter but different focal ratio, would produce the same image spreads, measured in seconds of arc,

§2. THE PLATE-PROFILE EXPANSION

Following the notation of the previous papers, we choose as our unit of length the radius of the primary mirror R and denote by ξ the ratio of the radius of the secondary mirror to that of the primary. Thus, if f_1, f_2 denote the paraxial focal lengths of the primary and secondary mirrors respectively,

$$f_1 = \frac{1}{2}, \quad f_2 = \xi/2 \quad \text{with} \quad R = 1.$$

We write $\mu = H/R$ where H is the radius of the aperture stop, and suppose that μ^2 is small compared with unity. For example, in the case of an $f/3.5$ system $\mu^2 = 1/36$, very nearly. Consequently the occurrence of an error-factor $[1 + O(\mu^2)]$ in an equation implies an inaccuracy of only a few per cent.

If we set up coordinate axes (x, y) with centre C (Figure 1) in the aperture stop and write $x^2 + y^2 = H^2 r^2$, then r denotes any zone-radius in the aperture and runs from 0 at the centre to 1 at the marginal zone.

Then if $T(r) \equiv T(x, y)$ denotes the plate thickness we can write the profile of the corrector plate in the form,

$$S(r) = \frac{T(r) - T(0)}{R} = a_1 \mu^2 r^2 + a_2 \mu^4 r^4 + a_3 \mu^6 r^6 + O(\mu^8). \quad \dots\dots(1)$$

If, in (1), we take $a_1 = 0$, a_2 and a_3 can still be chosen to give a plate which produces axial stigmatism in the light of one wavelength to the same degree of approximation. This first term in the expansion is introduced in order that the steepest plate slope occurring over the aperture may be reduced. This facilitates figuring, and in practice a_1 may be chosen to lessen both monochromatic and chromatic aberrations. Because of this term the plate possesses a positive paraxial focal length f_p given by

$$f_p = -\frac{1}{2(n_0 - 1)} \cdot \frac{1}{a_1}, \quad \dots\dots(2)$$

where n_0 is the refractive index of the corrector plate.

In the equation of the plate profile the steepest plate slope can be minimized by equating the slope at the marginal zone to the maximum slope occurring over the rest of the aperture. In between there is a 'neutral zone' whose radius will be denoted by $r = r_0$, at which the plate slope is zero and through which the rays pass undeflected.

In Figure 1 a ray QP_1 parallel to the axis of the system is shown to pass through the neutral zone at Q and, after reflection at the two mirrors, to cut the axis in F . If, using Gaussian optics, we image F backwards through the system to infinity by adjusting f_p , we can find a relation between f_p and r_0 which expresses the fact that F is then the paraxial focus of the system.

In Figure 1

$$\sin \theta_0 = \mu r_0 / R = \mu r_0 \quad \dots\dots(3)$$

and

$$CL = \mu r_0 \operatorname{cosec} 2\theta_0 \quad \text{and} \quad CP_2 = \xi.$$

Therefore, from the triangle CLP_2 ,

$$\frac{\sin \psi_0}{\sin 2\theta_0} = \frac{\mu r_0 \operatorname{cosec} 2\theta_0}{\xi},$$

so that

$$\sin \psi_0 = \mu r_0 / \xi. \quad \dots\dots(4)$$

An application of the sine rule to triangle CFP_2 gives $CF = \xi \sin \psi_0 / \sin 2(\psi_0 - \theta_0)$, and thus

$$BF = \xi [\sin \psi_0 / \sin 2(\psi_0 - \theta_0)] - \xi. \quad \dots\dots(5)$$

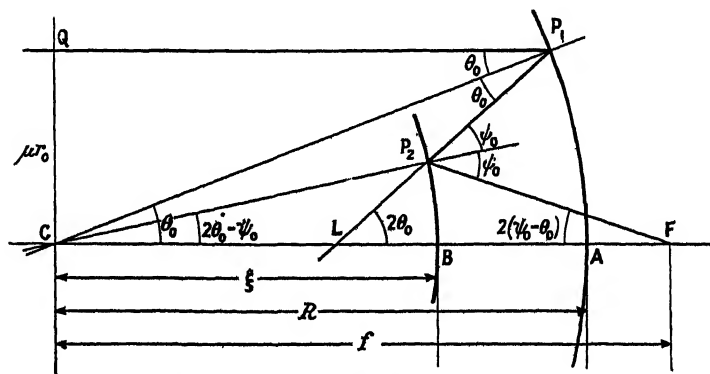


Figure 1.

The radius, $R=AC$, of the primary mirror is taken as unit of length.

If now we find G , the Gaussian image of F after reflection at the two mirrors, by making $f_p = 1 + AG$, we image G through the corrector plate paraxially to infinity. Using (5) we obtain

$$\frac{1}{f_p} = 2 \frac{\xi - 1}{\xi} + \frac{1}{\xi} \frac{\sin 2(\psi_0 - \theta_0)}{\sin \psi_0}. \quad \dots\dots(6)$$

ψ_0 and θ_0 are directly related to r_0 by (3) and (4), and so we may write (6) as

$$\frac{1}{f_p} = A(r_0). \quad \dots\dots(7)$$

Equations (2) and (7) relate a_1 and r_0 . We now obtain two more equations involving a_1, a_2, a_3 and r_0 , enabling us to find expressions for these three plate-profile coefficients in terms of the parameter r_0 of the neutral zone.

We make use of the two following facts to obtain the required equations: (a) that the plate slope vanishes at $r=r_0$; (b) that under perfect axial stigmatism the optical paths from a wave-front entering the system to the focus is the same along the axial ray and along the ray through the neutral zone.

Because of (a) we have, by differentiating (1),

$$2a_1\mu^2r_0 + 4a_2\mu^4r_0^3 + 6a_3\mu^6r_0^5 = O(\mu^8)$$

or
$$2a_2\mu^4r_0^4 + 3a_3\mu^6r_0^6 = -a_1\mu^2r_0^2 + O(\mu^8). \quad \dots\dots(8)$$

Because of (b) the path lengths $FP_2 + P_2P_1 + P_1Q$ and $FB + BA + AC$ must differ by an amount equal to the difference of retardation imposed by the plate, parallel to the axis, at the neutral zone and at the pole, i.e.

$$(FP_2 + P_2P_1 + P_1Q) - (FB + BA + AC) = -(n_0 - 1)S(r_0). \quad \dots\dots(9)$$

From Figure 1, $FP_2 = \xi \sin(2\theta_0 - \psi_0) / \sin 2(\psi_0 - \theta_0),$

$$LP_2 = \xi \sin(2\theta_0 - \psi_0) / \sin 2\theta_0,$$

and $LP_1 = \mu r_0 \operatorname{cosec} 2\theta_0 = \frac{1}{2} \sec \theta_0.$

Therefore $P_1P_2 = LP_1 - LP_2 = \frac{1}{2} \sec \theta_0 - \xi \sin(2\theta_0 - \psi_0) / \sin 2\theta_0.$

Also $P_1Q = \cos \theta_0,$

$$FB = \xi \left[\frac{\sin \psi_0}{\sin 2(\psi_0 - \theta_0)} - 1 \right],$$

$$BA = 1 - \xi; \quad AC = 1.$$

Then

$$\begin{aligned} & (FP_2 + P_2P_1 + P_1Q) - (FB + BA + AC) \\ &= \xi \left[\frac{\sin(2\theta_0 - \psi_0) - \sin \psi_0}{\sin 2(\psi_0 - \theta_0)} - \frac{\sin(2\theta_0 - \psi_0)}{\sin 2\theta_0} + 2 \right] + \frac{1}{2} \sec \theta_0 + \cos \theta_0 - 2 \\ &= B(r_0), \text{ say.} \end{aligned} \quad \dots\dots(10)$$

Then, from (9) and (10),

$$a_1\mu^2r_0^2 + a_2\mu^4r_0^4 + a_3\mu^6r_0^6 = -\frac{1}{n_0 - 1} B(r_0) + O(\mu^8)$$

or
$$a_2\mu^4r_0^4 + a_3\mu^6r_0^6 = -a_1\mu^2r_0^2 - \frac{1}{n_0 - 1} B(r_0) + O(\mu^8). \quad \dots\dots(11)$$

Using (2), (7), (8) and (11) we find

$$\left. \begin{aligned} a_1 &= \frac{-1}{n_0 - 1} A(r_0), \\ a_2\mu^4r_0^4 &= \frac{1}{n_0 - 1} A(r_0)\mu^2r_0^2 - \frac{3}{n_0 - 1} B(r_0), \\ a_3\mu^6r_0^6 &= \frac{-1}{2(n_0 - 1)} A(r_0)\mu^2r_0^2 + \frac{2}{n_0 - 1} B(r_0). \end{aligned} \right\} \quad \dots\dots(12)$$

We now need power series expansions for $A(r_0)$ and $B(r_0)$.

From (3) and (4) $\sin \theta_0 = \mu r_0$ and $\sin \psi_0 = \mu r_0 / \xi$, whence we can obtain

$$\left. \begin{aligned} A(r_0) &= \frac{\xi^3 - 4\xi^2 + 4\xi - 1}{\xi^3} \mu^2r_0^2 + \frac{\xi^5 - 8\xi^3 + 8\xi^2 - 1}{4\xi^5} \mu^4r_0^4 + O(\mu^6), \\ B(r_0) &= \frac{\xi^3 - 4\xi^2 + 4\xi - 1}{\xi^3} \mu^4r_0^4 + \frac{\xi^3 - 4\xi^4 + 6\xi^3 - 6\xi^2 + 4\xi - 1}{4\xi^5} \mu^6r_0^6 + O(\mu^8). \end{aligned} \right\} \quad \dots\dots(13)$$

Substituting into (12) we finally obtain for the plate-profile coefficients:

$$\left. \begin{aligned} (n_0-1)a_1 &= -\frac{\xi^3-4\xi^2+4\xi-1}{2\xi^3}\mu^2r_0^2-\frac{\xi^5-8\xi^3+8\xi^2-1}{8\xi^5}\mu^4r_0^4+O(\mu^6), \\ (n_0-1)a_2 &= \frac{\xi^3-4\xi^2+4\xi-1}{4\xi^3}-\frac{\xi^5-6\xi^4+13\xi^3-13\xi^2+6\xi-1}{2\xi^5}\mu^2r_0^2+O(\mu^4), \\ (n_0-1)a_3 &= \frac{3\xi^5-16\xi^4+32\xi^3-32\xi^2+16\xi-3}{8\xi^5}+O(\mu^2). \end{aligned} \right\} \dots\dots(14)$$

Combining (14) with (1), the plate-profile expansion can be written as *

$$S(r) = \frac{\xi^3-4\xi^2+4\xi-1}{4(n_0-1)\xi^3}\mu^4(r^4-ar^2)+O(\mu^6), \quad \dots\dots(15)$$

where $a=2r_0^2$.

If we take $a=3/2$, $\partial^2S/\partial r^2=0$ at $r=\frac{1}{2}$ and then

$$\left|\frac{\partial S}{\partial r}\right|_{r=1} = \left|\frac{\partial S}{\partial r}\right|_{r=1/2}$$

so that the maximum plate slope occurring over the aperture is minimized. This is the condition for minimum axial colour spread (Stroemgren 1935). The radius of the neutral zone is then given by, $r_0=\sqrt{3/2}\simeq 0.866$, and the shape of the plate profile is the same as in the colour-minimized classical Schmidt camera.

If, using the sixth-power terms in the plate-profile expansion as well, a more exact value of r_0 for axial colour minimization is obtained, the departure from the $\sqrt{3/2}$ value is negligible even in wide-aperture systems ($\mu=\frac{1}{2}$), and no appreciable improvement in performance could result.

The focal length of the system, represented by CF in Figure 1, is given by

$$\begin{aligned} CF &= \xi \sin \psi_0 / \sin 2(\psi_0 - \theta_0) = \mu r_0 \operatorname{cosec} 2(\psi_0 - \theta_0) \\ &= \frac{\xi}{2(1-\xi)} \left\{ 1 + \frac{\xi^2-3\xi+1}{2\xi^2}\mu^2r_0^2 + \frac{3\xi^4-11\xi^3+15\xi^2-11\xi+3}{8\xi^4}\mu^4r_0^4 + O(\mu^6) \right\}. \end{aligned} \quad \dots\dots(16)$$

The nominal aperture ratio of the system is therefore

$$F=f/2\mu = \frac{\xi}{4\mu(1-\xi)} \left\{ 1 + \frac{\xi^2-3\xi+1}{\xi^2}\frac{3}{8}\mu^2 + O(\mu^4) \right\}$$

$$\text{and} \quad \lambda = \frac{1}{4F} = \mu \frac{(1-\xi)}{\xi} \left\{ 1 - \frac{\xi^2-3\xi+1}{\xi^2}\frac{3}{8}\mu^2 + O(\mu^4) \right\}, \quad \left\{ \right.$$

$$\text{while conversely,} \quad \mu = \lambda \frac{\xi}{(1-\xi)} \left\{ 1 + \frac{\xi^2-3\xi+1}{\xi^2}\frac{3}{8}\lambda^2 + O(\mu^4) \right\}. \quad \left. \right\} \dots\dots(17)$$

§ 3. THE ABERRATION-DISPLACEMENTS

Using the argument of Hawkins and Linfoot (1945), we can, as explained above, obtain from equation (3.4) of that paper an expression for the leading monochromatic terms in the components of angular aberration for a monocentric system:

$$\delta X + i\delta Y = (n-1)\sin^2\phi \left(\frac{n+1}{2n}\frac{\partial T}{\partial x} + \frac{x}{2}\frac{\partial^2 T}{\partial x^2} + \frac{i}{2n}\frac{\partial T}{\partial y} + \frac{ix}{2}\frac{\partial^2 T}{\partial x\partial y} \right). \quad \dots\dots(18)$$

* The 'plate-strength' is given by the term in r^4 in equation (15). This result agrees with equation (15) in Linfoot (1944), obtained by plate diagram analysis.

As before, $T(x, y)$ is the thickness of the corrector plate which has refractive index n . ϕ is the off-axis angle of an image point in the plane containing Cx and the axis of the system.

We take coordinates (u, v) in the aspheric surface such that $x = \mu u R$, $y = \mu v R$. Then u and v each lie in the range $[-1, +1]$ and the marginal zone of the corrector plate is represented by $r^2 \equiv u^2 + v^2 = 1$. To keep the obstruction caused by the secondary mirror, which is partly dependent on ϕ , reasonably small, we must have the semi-angle ϕ_0 of the order of μ .

Now $T(x, y) - T(0, 0) = SR$, so that $\frac{\partial T}{\partial x} = \frac{1}{\mu} \frac{\partial S}{\partial v}$, etc.

Then from (15) and (18) we find apparent angular aberrations $(\delta X, \delta Y)$, in seconds of arc, to be

$$\delta X + i\delta Y = \frac{\xi^3 - 4\xi^2 + 4\xi - 1}{4\xi^3} K \mu^3 \phi^2 \left[\frac{n_0 + 1}{2n_0} \frac{\partial}{\partial \mu} + \frac{1}{2} u \frac{\partial^2}{\partial u^2} + \frac{i}{2n_0} \frac{\partial}{\partial v} + \frac{1}{2} i u \frac{\partial^2}{\partial u \partial v} \right] (r^4 - ar^2) + O(K\mu^7), \quad \dots (19)$$

where $K = 648,000/\pi$.

Effects of plate tilt other than those represented in (18), together with the effect of the terms of the order of μ^6 in the plate-profile expansion, are included in the error term $O(K\mu^7)$ in (19). This equation then represents the leading monochromatic aberrations of the monocentric Schmidt-Cassegrain camera, and these are seen to be dependent, to a first approximation, only on the second- and fourth-power terms of the plate-profile expansion.

In practice the corrector plate is usually figured to give as good axial stigmatism as possible under an autocollimation test. Before polishing, the plate is ground to the profile given by equation (1), which represents this final form with the stated accuracy. The sixth-power terms are included in the explicit formulae because in wide-aperture systems their neglect may cause errors in the figuring depth greater than can easily be corrected in polishing (see Linfoot 1948, p. 285).

If we take n_0 to be the refractive index for the light in which the system is axially stigmatic, equation (19) represents the aberrations in light of that wavelength completely. In light of other wavelengths the refractive index of the plate is in practice $n = n_0 + O(\mu^2)$.

We can write the chromatic aberrations as the angular change with wavelength in deviation by the plate. This deviation, for any ray meeting the plate in a point (u, v) , inclined to the axis at an angle not greater than the order of μ , is given by *

$$\frac{n-1}{\mu} \left(\frac{\partial S}{\partial u} + i \frac{\partial S}{\partial v} \right) (1 + O(\mu^2)).$$

Therefore we can write, for the angular chromatic aberrations, in seconds of arc,

$$\begin{aligned} \delta X + i\delta Y &= \frac{K}{\mu} (n - n_0) \left(\frac{\partial S}{\partial u} + i \frac{\partial S}{\partial v} \right) (1 + O(\mu^2)) \\ &= \frac{\xi^3 - 4\xi^2 + 4\xi - 1}{4\xi^3} K \mu^3 \frac{n - n_0}{n_0 - 1} \left(\frac{\partial}{\partial u} + i \frac{\partial}{\partial v} \right) (r^4 - ar^2) (1 + O(\mu^2)). \end{aligned} \quad \dots (20)$$

Since $n - n_0$ is of the order of μ^2 , the error term in (20) is

$$O(K\mu^3)O(\mu^2)O(\mu^2) = O(K\mu^7),$$

* See Hawkins and Linfoot (1945) equation (3.1).

and thus (20) represents the chromatic aberrations to the same degree of accuracy as (19) represents the monochromatic aberrations.*

The total components of angular aberration for an off-axis image point in light for which the refractive index of the corrector plate is n are represented by the addition of (19) and (20), the cross terms involved being of the order of $K\mu^7$.

§ 4. A TYPICAL SYSTEM

Choice of a value for the parameter ξ is restricted to a small range, for while it is desirable that it should be as low as possible, so that the obstruction caused by the secondary mirror shall not be too great, the focal surface must lie behind the primary mirror if we are to obtain the advantage of a fully accessible image-surface which a Schmidt-Cassegrain system permits. Thus we must have f greater than 1 or, from (17), to a first approximation, $\xi/2(1-\xi) > 1$, i.e. $\xi > \frac{2}{3}$.

Allowing for the thickness of the mirror, $\xi = 0.69$ seems to be the lowest permissible value.

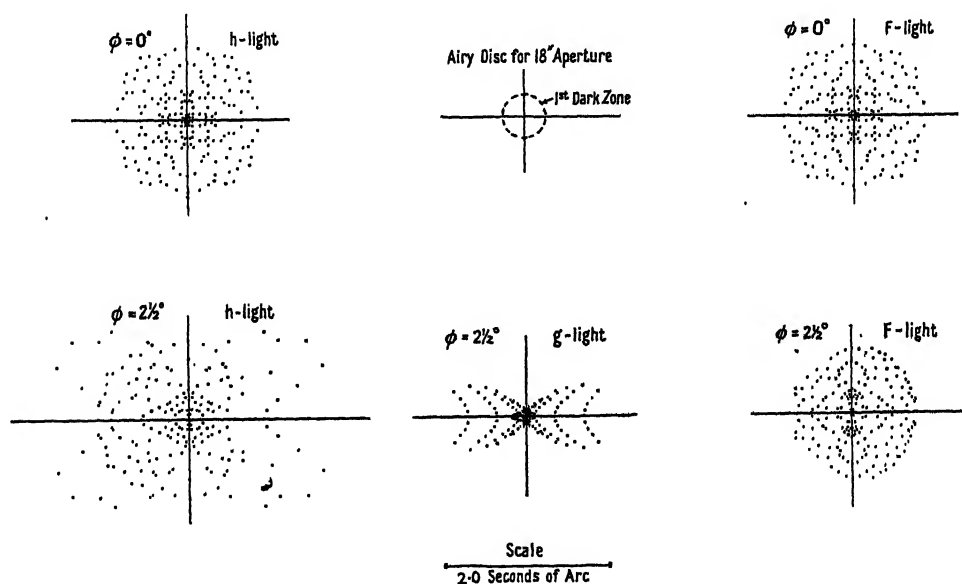


Figure 2. Spot-diagrams of the image-points of an $f/3.5$ monocentric Schmidt-Cassegrain camera at the centre and edge of a five-degree field in F, g and h-light and the Airy disc of an 18-inch system.

Because of the curvature of the image-surface, the system cannot conveniently be used over a greater field diameter than five degrees, corresponding to $\phi_0 = 2\frac{1}{2}^\circ = 0.0436$ radian. For such off-axis angles the chromatic aberrations are more important than the monochromatic aberrations, and if, over a spectral range F-h, which includes the brightest part of an ordinary photographic spectrum, the geometrical images are not to have an effective diameter greater than about two seconds of arc, the lowest nominal aperture-ratio permissible is about $f = 3.5$.

Figure 2 shows spot diagrams of the geometrical images of an axially colour-minimized $f/3.5$ system at the centre and at the edge of a five-degree field, in F, g and h light. Each diagram is obtained by plotting as spots the $(\delta X, \delta Y)$ which

* It is evident that the diameter of the chromatic disc in the light of any wavelength is directly proportional to the change in the refractive index of the corrector plate.

correspond to the rays from a single object point passing through the points $u, v = 0, \pm 0.1, \pm 0.2, \dots$; $u^2 + v^2 \leq 1$ of a lattice of small squares filling the entry-pupil. The spot density may be taken as an indication of the light intensity in the aberrated image. Rays which are intercepted by the secondary mirror have been excluded; hence there is asymmetry in the off-axis image points. A quantitative picture of the part played by diffraction in image formation is not easily to be obtained from the size of the Airy disc (see Figure 2) because the wave fronts entering the system suffer continual small distortions due to atmospheric disturbances. Under average 'seeing' conditions the size of this 'tremor disc' will be as much as two seconds of arc in diameter.

It seems that the ordinary Schmidt system, which might best be compared with the above system, would be that having the same diameter of aperture and a

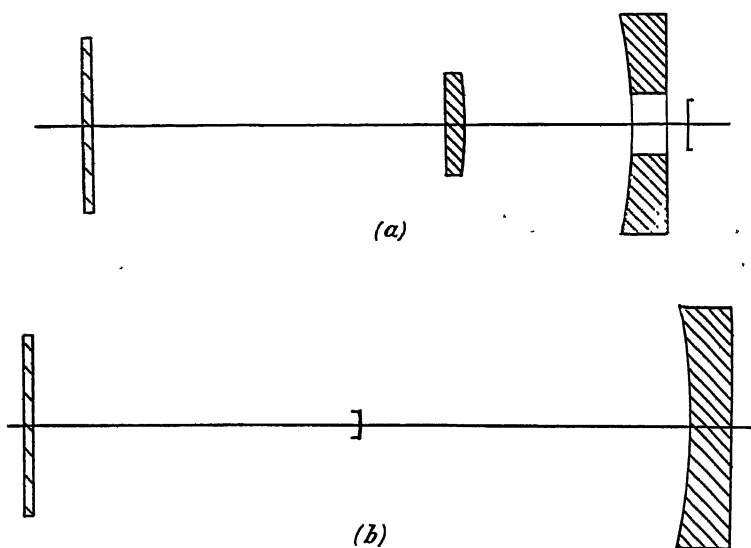


Figure 3.

(a) $f/3.5$ monocentric Schmidt-Cassegrain camera.

(b) $f/1.9$ ordinary Schmidt camera with the same aperture diameter.

focal ratio of $f/1.9$. The depth of figuring on the plate would then, to a first approximation, be the same as on the $f/3.5$ monocentric Schmidt-Cassegrain system. Because of this similarity the two systems will, over a five-degree field, have practically the same angular aberrations. Moreover, because their apertures are of equal size, the angular diameter of the diffraction discs will be the same in the two systems. Both from the geometrical standpoint, therefore, and from the more precise diffraction theory of image formation, a given star-field will be rendered in very much the same way in the two systems but on a smaller plate and with greater speed on the $f/1.9$ ordinary Schmidt camera. If allowance is made for the smaller light loss within the latter system, partly through the reduced obstruction and partly through the elimination of a second reflection, the difference in exposure times with the same aperture diameter might amount to a factor of five-thirds.

The larger plates used on the Schmidt-Cassegrain system would exhibit the defects of grain size and sky-blackening to a lesser extent than would the plates of half the size used on the ordinary Schmidt camera. This would be an advantage of some importance except in systems where the focal length is already greater than about 100 inches. Figure 3 shows the two systems drawn to the same scale; their external dimensions are seen to be comparable.

ACKNOWLEDGMENTS

The author wishes to express his gratitude to Dr. E. H. Linfoot and Dr. E. Wolf for their interest and encouragement in the preparation of this paper, and to the former for the loan of his paper (1949) in advance of publication.

REFERENCES

- BAKER, J. G., 1940, *J. Amer. Phil. Soc.*, **83**, 3, 339.
 BURCH, C. R., 1942, *Mon. Not. R. Astr. Soc.*, **102**, 159.
 CARATHÉODORY, C., 1940, *Hamburger Mathematische Einzelschrift*, **28**, 36.
 HAWKINS, D. G., and LINFOOT, E. H., 1945, *Mon. Not. R. Astr. Soc.*, **105**, 334.
 LINFOOT, E. H., 1944, *Mon. Not. R. Astr. Soc.*, **104**, 48; 1949, *Ibid.*, **109**, 279.
 STROEMGREN, B., 1935, *Vierteljahresschrift der Astr. Ges.*, **70**, 65.

The Operation and Photographic Characteristics of a Kerr-cell Type of Electro-optical Shutter

By A. E. J. HOLTHAM AND H. A. PRIME

Electronics Department, University of Liverpool

MS. received 13th December 1949

ABSTRACT. A Kerr-cell electro-optical shutter is described, the shutter action being controlled by a high voltage pulse applied to the cell electrodes. The exposure time used is 2×10^{-6} second, and provision is made for obtaining 'delayed' exposures up to times of the order of 10^{-4} second. The photographic limitations of the shutter are discussed and the formation of the image considered in some detail. The assumptions necessary for a quantitative photometric analysis of Kerr-cell exposures are stated and an approximate limit is set on the magnitude of errors resulting from such assumptions. The applications of the Kerr-cell technique to the measurement of particular spark channel characteristics are also described.

§ 1. INTRODUCTION

CERTAIN isotropic liquids, such as nitrobenzene, become doubly refracting when subjected to a high electrical stress. This property, first observed by Kerr in 1875, and subsequently known as the 'Kerr effect', has been used by various investigators in the development of electrically controlled light shutters, by which the transmission of light can be controlled down to times of the order of 10^{-8} second or less. The time lag between the occurrence of the 'Kerr effect', and the application of the stress (to two electrodes immersed in the nitrobenzene or similar liquid) is not greater than 10^{-8} second, and hence the rate of opening or closing the shutter depends mainly on the rate of rise or fall of the voltage applied to the electrodes.

Some workers (Abraham and Lemoine 1900, Dunnington and Lawrence 1930, Dunnington 1931) have used the 'open-to-closed' type of shutter, in which the shutter, a Kerr cell placed between 'crossed' polarizing Nicol prisms, is initially kept open by the application of a steady D.C. voltage to the cell plates. The shutter is then closed by the removal of the applied voltage at some instant subsequent to the initiation of the event being observed. A 'closed-to-open-to-closed' shutter, employing two cells between crossed Nicols, has also been developed (Beams 1926), in which both cells are initially under stress, the directions of the two electric fields being mutually perpendicular, thus giving an optically 'closed' shutter. To obtain the exposure, the voltage on one cell is removed an instant before that on the other. This permits a light pulse to be transmitted, the intensity depending on the difference in the voltages on the two cells at any given instant.

An alternative method of obtaining a 'closed-to-open-to-closed' shutter action, using only a single cell, is by the application of a voltage pulse to the cell electrodes. The duration of the pulse determines the exposure time, and the steepness of the leading and trailing edges of the pulse determines the rates of opening and closing of the shutter. An advantage of this technique is that the pulse may be applied at any instant relative to the initiation of the phenomenon, but a limitation is that the exposure times are not as short as those obtainable with the previous techniques. Nevertheless, the method appears to be the one mainly used at the present time (Lapsley, Snoddy and Beams 1948, Rawcliffe 1942, Froome 1948, Zarem, Marshall and Poole 1949). A recent feature of these single cell pulse techniques is the use of Polaroid film in place of Nicol prisms as the polarizing elements in order to improve the aperture of the optical system. A description of a Kerr-cell equipment based on this technique is given in §2 of the paper.

Although several of the papers quoted, and also others, consider the electro-optical characteristics of the various types of shutter in some detail (Dunnington 1931, Washburn 1932, Zarem *et al.* 1949), there has been practically no discussion of the photographic characteristics of the shutter from the point of view of image formation and the photometric analysis of results obtained. This aspect of the shutter mechanism is considered in §3 of the paper. Since the equipment described has been primarily used for investigations concerned with the growth of a spark channel, the results quoted in the text refer to this type of optical transient (Prime and Saxe 1949). In the final section of the paper consideration is given to the application of the technique to the determination of the channel diameter and the variation of radial intensity across a spark.

§ 2. DESCRIPTION OF THE EQUIPMENT

(i) *Electronic Equipment*

A schematic diagram of the apparatus is shown in Figure 1. A hydrogen thyatron modulator circuit is used to generate a pulse, the amplitude of which can be adjusted by varying the H.T. supply voltage to the hydrogen thyatron from almost zero to 4 kv. The pulse is subsequently stepped up to a maximum of 10 kv. by means of a pulse transformer for direct application to the cell electrodes. The duration of the pulse is determined by the characteristics of the artificial line forming the pulse, and is approximately $2\mu\text{sec}$. A trigger circuit, for tripping the hydrogen thyatron at any time relative to the initiation of the event under

observation, and associated power units are contained in a single unit with the modulator. In addition, a 3-in. cathode-ray tube is incorporated for monitoring the performance of the equipment.

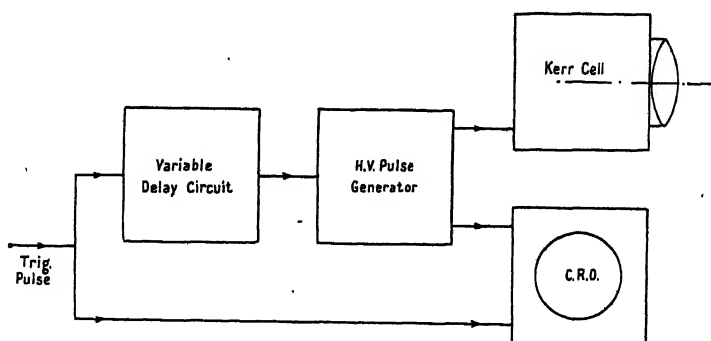


Figure 1. Schematic diagram of apparatus.

A diagram of the trigger, delay and modulator circuits is shown in Figure 2. Facilities are provided for triggering the system by means of a positive or negative external pulse or by a recurrent pulse generated internally in the multivibrator stage V_1 . V_2 serves to reverse the phase of the external negative pulse so that a positive pulse is always applied to the first grid of V_3 . This positive pulse can be monitored on the oscillograph. The switch S_1 controls the operation of V_1 ; V_3 and V_4 comprise the delay stages enabling a variable delay from 0–250 μ sec.

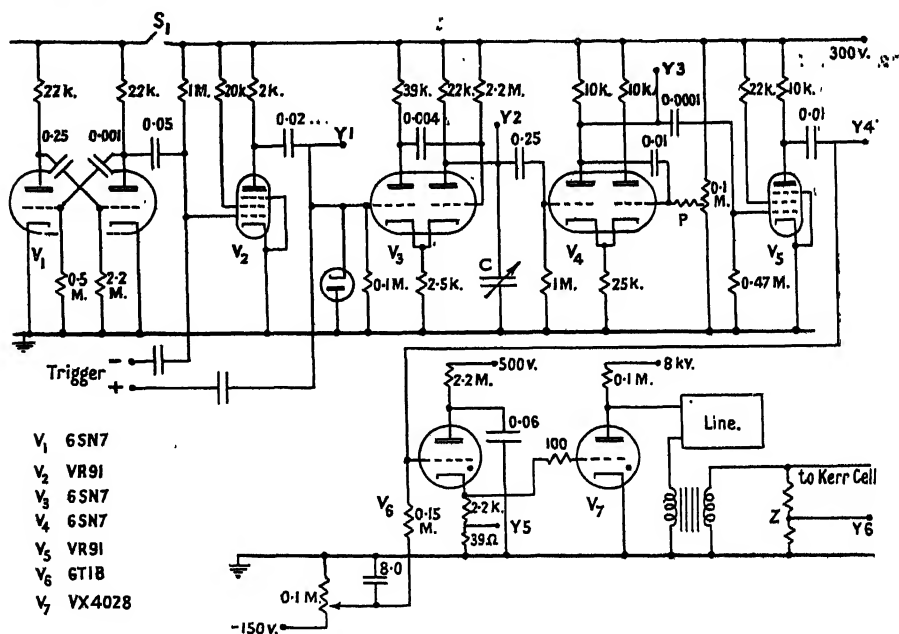


Figure 2. Circuit diagram of delay and pulsed modulator unit.

to be obtained. The output from the right-hand anode of V_3 is a positive-going saw-tooth wave which is fed to the grid of the voltage comparator stage V_4 . The rate of rise of the saw-tooth can be controlled by variation of the capacitor C ,

between the output anode and earth. The comparator stage V_4 operates when the instantaneous voltage of the saw-tooth is equal to the potential set by the potentiometer connected to the right-hand grid. Variation of C, by switching a bank of condensers, provides a coarse delay control while variation of the potentiometer P enables a fine control of delay to be obtained on any given range. The output from the comparator stage is a negative square pulse generated at the left-hand anode. The pulse is amplified, and the phase reversed in the following valve V_5 which, in turn, triggers a small argon thyatron V_6 . The hydrogen thyatron V_7 requires to be triggered from a low impedance source and by a positive-going pulse of the order of 100–200 volts amplitude; the use of an argon thyatron between V_6 and V_7 fulfils these requirements.

A hydrogen thyatron was chosen as the modulator valve for the following reasons: (i) The trigger-pulse amplitude is very much less than would be required by a triggered spark gap or trigatron type of valve. (ii) The current-handling capacity (~ 90 amp.) greatly exceeds that of a hard valve. In addition, the operation of the thyatron is independent of the shape of the trigger pulse subsequent to the valve 'firing'. (iii) The hydrogen thyatron is not temperature dependent as would be a mercury thyatron, and the ionization and de-ionization times are very much less than would obtain for a mercury thyatron, due to the higher ion mobility. Also, the hydrogen thyatron has a positive grid control characteristic which avoids the necessity of a bias supply for the grid circuit.

A detailed description of the operation of hydrogen thyatrons has been given by Germeshausen (1948), and it has been shown that the delay between the application of the grid pulse and the 'firing' of the anode circuit, is of the order of $0.2 \mu\text{sec.}$ for a rate of rise of grid voltage of $1,500 \text{ volts}/\mu\text{sec.}$ This condition is realized in the equipment described above.

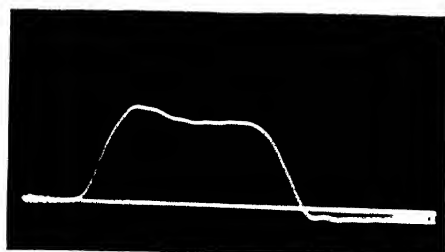
The operation of the modulator circuit itself is conventional, the characteristic impedance of the artificial line being matched to the load on the secondary side of the pulse transformer.

Points indicated Y1–Y5 on the circuit diagram can be monitored by the oscillograph, the circuit diagram of which is shown in Figure 3. The wave-forms shown in Figure 5 (photograph) show the shape of the pulse applied to the Kerr-cell electrodes, and were obtained using the internal trigger pulse and a delay of the order of $\frac{1}{2} \mu\text{sec.}$ The delay is constant from pulse to pulse within approximately 10^{-7} sec. , thus permitting a steady oscillogram to be obtained under recurrent pulse conditions.

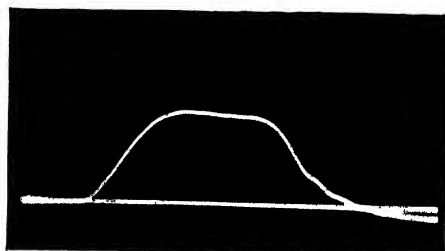
(ii) *Optical Equipment and Method of Adjustment*

An aerial camera was modified to take the two Polaroid discs and the cell. The lens of the camera is a Kodak 'Ektar' $f/2.5$ of 7 in. focal length. The polarizer is fixed into the focal-plane shutter compartment, and can be rotated by a reduction gearing. An extension has been fitted to the back of the camera and contains the cell and analyser Polaroid, which can also be rotated. A 35-mm. film adaptor can be attached to the end cover of the extension for recording purposes. An overall magnification of unity is obtained with the above system.

The preparation of the copper electrodes for the Kerr cell was carried out in the following manner: the surfaces were first milled, then rubbed with emery cloth, cleaned and finally coated with black cupric oxide in order to minimize internal reflections from the electrodes. The oxide deposit was obtained

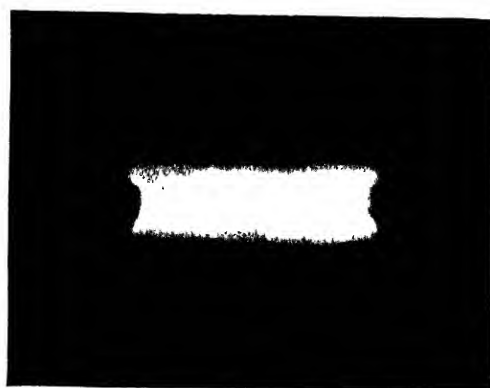


(a) Cell unconnected.



(b) Cell connected to modulator.

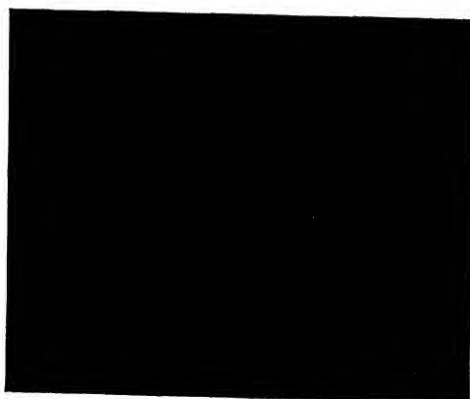
Figure 5. Oscillograms of Kerr-cell exposure pulse.



(a)



(b)



(c)

Figure 6. Series of Kerr-cell records for 2,000-amp. spark in air at atmospheric pressure. Exposure time 2 μ sec. at (a) 0-2 μ sec.; (b) 3-5 μ sec.; (c) 5.5-7.5 μ sec. Duration of current pulse 15 μ sec.

electrolytically in a chemically normal solution of caustic soda as the electrolyte. The size of the electrodes and their spacing is arranged to give maximum light transmission for a 10 kv. pulse; the length of the electrodes along the optic axis is 1.5 cm., the separation 0.34 cm., and the height 5 cm. The nitrobenzene is distilled into the cell, which is then sealed using water-glass as a cement.

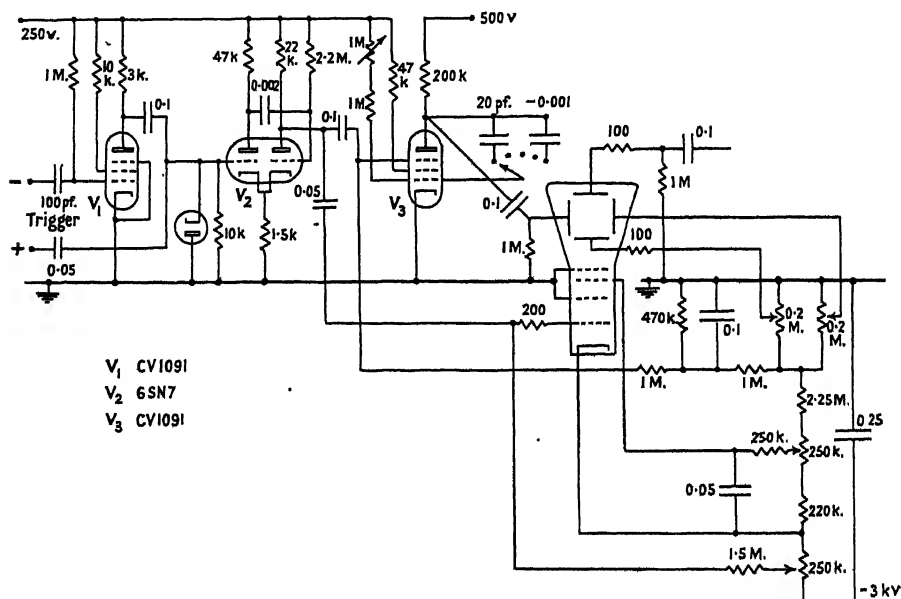


Figure 3. Circuit diagram of cathode-ray oscillograph monitor unit.

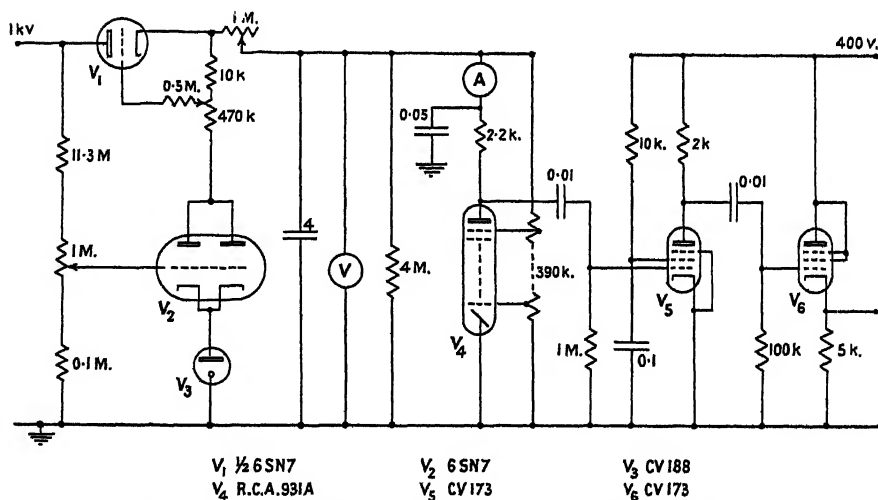


Figure 4. Circuit diagram of photo-multiplier and amplifier unit.

In order to adjust and check the performance of the equipment, a photomultiplier cell and amplifier together with the necessary rectifier circuits was developed for use with the monitoring oscilloscope to record the light transmitted through the Kerr-cell unit. The circuit diagram of this unit is shown in Figure 4.

For optimum conditions of light transmission, the polarizer and analyser should be crossed and their planes of polarization should be at 45° to the plane of the cell electrodes (Kingsbury 1930). This adjustment is made using a steady light source. For a given orientation of the analyser, the polarizer is slowly rotated by the reduction gearing until the light transmission to the photo-multiplier tube is a minimum. This position is indicated by a minimum reading on the milliammeter in series with the photo-multiplier tube.

A recurrent voltage pulse is next applied to the Kerr cell and the transmitted light pulse picked up by the photo-multiplier tube and recorded on the oscillograph screen. The analyser and polarizer are then rotated together until the amplitude of the pulses is a maximum. Finally, a small correction of the polarizer may be necessary to ensure that the polarizer and analyser are still crossed.

Even under conditions of optimum adjustment, the overall transmission of the system is only of the order of a few percent due to the attenuation of light by the polarizer, nitrobenzene and the analyser. For example, Zarem *et al.* (1949) give the transmission of 'J' film Polaroid within the pass-band of the nitrobenzene as 32%, and as the Kerr cell itself has a theoretical transmission of only 50% when the Kerr effect is induced, the maximum overall transmission for the system is given approximately by $\frac{1}{3} \times \frac{1}{2} \times \frac{1}{3}$ or 5%.

§ 3. PHOTOGRAPHIC CHARACTERISTICS OF THE KERR-CELL SHUTTER

The equipment described in § 2 has been used to study the variation of the visible spark channel diameter, the variation of the maximum intensity of the channel and the variation of light intensity across the channel, as functions of time, gas, pressure and current. A typical series of results is shown in Figure 6. These were obtained for a 15 kv. spark in air, generated by the discharge of an artificial line through its characteristic impedance to give 2,000 amp. for 15 μ sec. The results of this work will be described in a separate publication, but it is of importance from the point of view of the general technique, to consider in some detail those factors that affect the interpretation and analysis of photographs obtained with the Kerr-cell shutter.

(i) *Exposure Density Characteristic and Film Calibration*

For any quantitative analysis of the Kerr-cell records, a knowledge is required of the exposure density characteristic of the film, which is normally expressed in terms of $\log E$ and D . It is desirable to use a film emulsion having a characteristic as linear as possible, and to restrict the exposure by means of an aperture so that the maximum density recorded is below the 'knee' of the characteristic.

A step-wedge calibration of the film can be made to determine the exposure density characteristic in terms of a relation between density and the relative transmission of each step of the wedge. The calibration is carried out under conditions which, as far as possible, are the same as those applying at the instant of normal spark recording. Thus a spark is used as the source of illumination in order to ensure that the spectral response of the emulsion introduces no additional complication, and the Kerr-cell exposure is maintained at the same value of 2 μ sec. The step-wedge is placed immediately in front of the film and a series of records is made of a given spark through successive steps of the wedge. Then, knowing the percentage transmission of each step of the wedge (see below), a

relation can be obtained between density (measured at a fixed arbitrary point on each record) and percentage transmission, taking the density recorded with no step-wedge to be 100%. By adjustment of the lens aperture, the maximum density recorded is arranged to be greater than that of any spark to be analysed.

The percentage transmission of the step-wedge must also be measured as far as possible under the same operating conditions in order to eliminate variations due to light scatter at the step-wedge surface. This is carried out by replacing the film by a photo-multiplier tube and using a uniformly illuminated slit at the focus of the lens, in place of the spark electrodes. By partial 'uncrossing' of the Polaroids, and adjustment of the voltage per stage of the photo-multiplier tube, a maximum anode current of about 250 μ a. for 100% transmission is obtained. For each step, the corresponding current is noted and hence the percentage transmission recorded. Within experimental error, a linear relation exists between light intensity and anode current up to 1 ma. (Engstrom 1947). Each individual reading has to be taken quickly to avoid fatigue of the tube and care has to be exercised to obtain steady D.C. potentials for both the photo-multiplier tube H.T. and the light source. This method does however neglect inter-reflections between the step-wedge and the photocell, which might be different from the inter-reflections between the step-wedge and the film, thus giving rise to an inaccurate transmission characteristic for the wedge.

(ii) *Formation of the Kerr-cell Image*

The Polaroids of the Kerr cell are imperfect in so far as a small percentage of the incident light is transmitted even when the two discs are 'crossed'. For the type of material used for this work, the manufacturers state that the degree of polarization for each element is approximately 99% (Polaroid Products Bulletin). This limitation results in the formation of a spurious image of the phenomenon on the photographic plate or film in the following manner. Assume that the duration of the optical transient being recorded is T sec. and that the duration of the Kerr-cell exposure is ΔT sec. and occurs at time t' ; in addition, denote the transmission of the cell when the Polaroids are 'crossed' by the fraction p , and the transmission during the exposure by q . Then, if the intensity of the phenomenon is I_t , a function of time, the three components contributing to the exposure of the film are given by the integrands

$$\int_0^{t'} I_t p \, dt, \quad \int_{t'}^{t'+\Delta T} I_t q \, dt, \quad \text{and} \quad \int_{t'+\Delta T}^T I_t p \, dt,$$

of which the second is the significant exposure, the first and third being unwanted.

The resultant exposure of the image may not necessarily be proportional to the sum of these integrands owing to the existence of reciprocity-law failure and the Clayden effect. The latter effect (Mees 1945) relates to the resultant density of an image produced by a very brief exposure to light of extremely high intensity and a subsequent exposure to light of moderate intensity. Under such conditions the separate exposures do not add in a simple manner. Failure of the reciprocity law is due to the fact that the density D depends not only on the exposure E (=intensity \times time), but also on the actual values of these two factors. Failure of the law has been investigated by Webb (1933) and Berg (1940, 1946). Berg claims in his second paper that reciprocity failure is negligible providing the times involved are less than 5×10^{-5} sec.

These photographic considerations lead to the conclusion that the density of the recorded or 'normal' image may not necessarily be given by a function of the form

$$D_N = f \left\{ \int_0^{t'} I_t p \, dt + \int_{t'}^{t' + \Delta T} I_t q \, dt + \int_{t' + \Delta T}^T I_t p \, dt \right\}, \quad \dots\dots (1)$$

where f is determined by the exposure-density characteristic of the film.

For the purpose of analysis however, it is convenient to assume that equation (1) is in fact, exact. Accordingly, by the definition of exposure, we have

$$\begin{aligned} E_N &= \int_0^{t'} I_t p \, dt + \int_{t'}^{t' + \Delta T} I_t q \, dt + \int_{t' + \Delta T}^T I_t p \, dt \\ &= \int_0^{t'} I_t p \, dt + \int_{t'}^{t' + \Delta T} I_t (q - p) \, dt + \int_{t'}^{t' + \Delta T} I_t p \, dt + \int_{t' + \Delta T}^T I_t p \, dt. \end{aligned}$$

Similarly, the spurious image exposure is defined by the equation

$$E_{sp} = \int_0^T I_t p \, dt = \int_0^{t'} I_t p \, dt + \int_{t'}^{t' + \Delta T} I_t p \, dt + \int_{t' + \Delta T}^T I_t p \, dt,$$

so that

$$\int_{\Delta T} I_t (q - p) \, dt = E_N - E_{sp}.$$

Further, if I_t is approximately constant at the value $I_{t'}$ during the interval ΔT , then

$$I_{t'} = \frac{E_N - E_{sp}}{(q - p) \Delta T}. \quad \dots\dots (2)$$

Now the units of exposure and intensity are arbitrary and since by the previous definition, E_{sp} is a constant for the event it is convenient to use the ratio, $E_{sp}/E_N = \phi$ and hence obtain a general solution

$$I_t = k \frac{1 - \phi_t}{\phi_t}, \quad \dots\dots (2a)$$

where k is some constant having the dimensions of intensity. The assumption that equation (1) is exact, which was necessary for the derivation of equation (2), is also implicit in the experimental determination of E_N and E_{sp} for the following two reasons. Firstly, the density transmission curve, referred to in the preceding section, is prepared under conditions almost similar to those obtaining during normal recording. There is, therefore, a spurious image contribution to the recorded density. The calibration curve so obtained is thus particular to the time t' at which the exposure is made and should strictly only be used to analyse normal recordings for which the exposure is likewise made at time t' . Secondly, the transmission scale of the calibration curve can only be interpreted as an exposure scale provided that there is no reciprocity-law failure. It is clear, therefore, that the use of a common calibration curve for the determination of E_N from D_N over all t' and the determination of E_{sp} from D_{sp} is only justified when equation (1) is exact.

A comparison of calibration curves obtained for different values of the exposure time t' showed that the spread of separation of these curves (which were obtained from the records on a number of films) was comparable with the spread due to processing the films separately. A separate experiment had shown that the

agreement between curves obtained from different films, but for constant t' , was of the order of 20%. This discrepancy due to processing errors therefore sets an approximate limit on the errors likely to result from the assumptions made in the previous analysis.

§ 4. THE EXPERIMENTAL DETERMINATION OF SPARK CHANNEL CHARACTERISTICS

(i) *Determination of Intensity as a Function of Time*

In order to illustrate the considerations of the preceding section, the determination of the variation of intensity as a function of time is described for the case of the 2,000 amp. air spark previously mentioned. A series of photographs are taken on a single film of: (a) the normal image for various delays from the time of initiation, (b) the normal image, at time $t=0$, through a step-wedge, and (c) the spurious image (i.e. no pulse applied to the Kerr cell). The lateral density profiles (i.e. variation of density across the images) are then determined using a microphotometer and the density-transmission calibration curve is constructed from the recorded images detailed under (b). The lateral exposure values are then tabulated for the records detailed (a) and (c) by reference to the calibration curve, and hence the ratio E_{sp}/E_N can be determined for corresponding points on the profiles. The table shows the values of E_N , ϕ_i and I_i in arbitrary units. tabulated as functions of t' , the mid-time of the exposure.

$t' (\times 10^{-8} \text{ sec.})$	1.5	5.5	7	10	12.5	16.5
E_N	48	16	6.2	3.4	1.8	1.5
$\phi_i (E_{sp}=0.6)$	0.013	0.037	0.093	0.18	0.33	0.40
I_i (arbitrary units)	76	26	9.8	4.6	2.0	1.5

It is to be noted that the spurious image contribution increases from approximately 1% of the total exposure at the initial stage, to approximately 30% at the end of the phenomenon. It is essential therefore to correct for the contribution at low intensities.

(ii) *Spark Channel Diameter and the Radial Variation of Intensity*

Knowing the spurious image contribution and the film characteristic, the intensity profile is readily determined from the density profile; it is therefore possible to investigate variations in spark channel 'diameter' as a function of the discharge parameters such as current, gas, gas pressure, etc. For this purpose some arbitrary definition of channel diameter must be proposed and it is convenient to use the distance between points on the intensity profile for which the intensity is half the maximum value. The definition is not altogether satisfactory as it depends on the recording sensitivity (i.e. the ratio of maximum recorded density to the density of the 'fog' which forms the background for the image). However, since the 'skirts' of the profile are relatively steep, the dependence of the diameter, as defined above, on the maximum density is not critical; in fact, a change in aperture (which is equivalent to a change in maximum recorded density) by a factor of 3 results in an apparent change in diameter of only 20%.

The intensity profile, from which the channel diameter is determined as described above, only shows the lateral variations of intensity. If it is assumed that the channel is of circular cross section and radius R , and if the axis of the

channel is taken along Oz , see Figure 7(a), then the lateral intensity profile shows the variation of intensity I_x , as a function of x , when the channel is viewed along a direction parallel to the plane Oyz . Now the intensity I_x is related to the more fundamental quantity, radial intensity I_r , which is the intensity at a point distance r from the Oz axis, by the relationship:

$$I_x = k \int_x^R r I_r (r^2 - x^2)^{-1/2} dr, \quad \dots\dots(3)$$

where k is a numerical constant which can be neglected since the units are arbitrary.

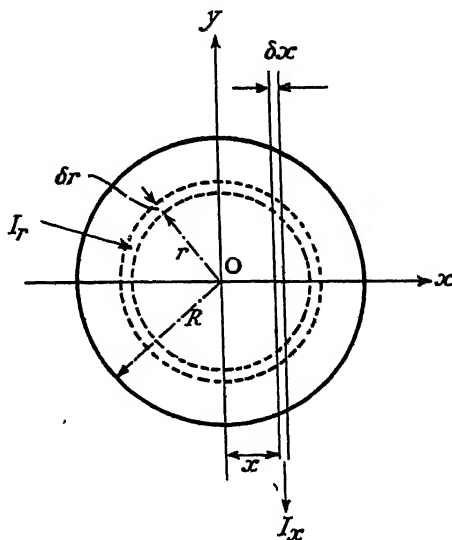


Figure 7(a). Relation between lateral intensity I_x and radial intensity I_r .

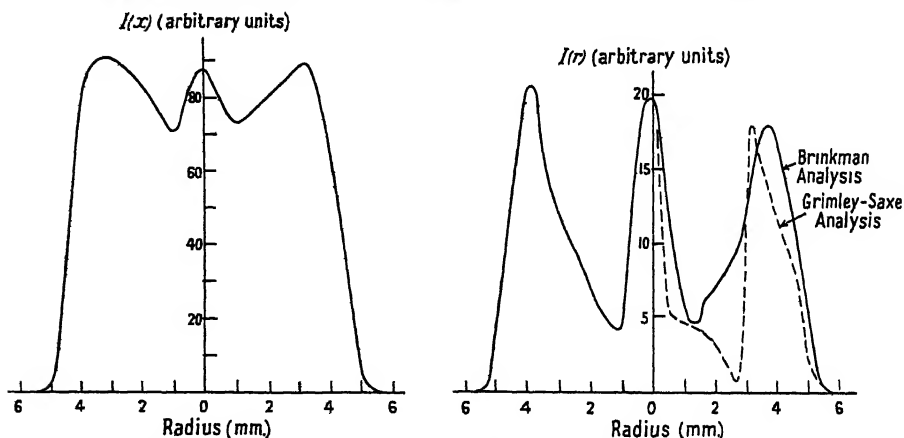


Figure 7(b). (i) Lateral intensity profile; (ii) radial intensity profile.

Solutions of equation (3), giving I_r as a function of I_x have been obtained by Hörmann (1935), Brinkman (1937) and Grimley and Saxe (1949). Brinkman obtains

$$I_r = \frac{1}{\pi} \frac{d}{dr} \int_r^R I'_x (x^2 - r^2)^{-1/2} dx \quad \dots\dots(4)$$

and an equivalent expression, in slightly different form, is obtained by Grimley and Saxe. Numerical integrations have been carried out to determine the radial intensity profiles for the case of a 2,000 amp. spark in hydrogen at a pressure of 1 atmosphere. For comparison the I_x profile is shown with the I_r profiles in Figure 7(b). The channel diameter as determined from the I_r profile according to the same definition is approximately 7% more than the diameter determined from the I_x profile.

§ 5. CONCLUSIONS

The conclusions that can be drawn from the work described in the paper relate to the photographic aspects of the shutter mechanism rather than to the electro-optical characteristics. Nevertheless, the technique described in §2 illustrates in some detail the operation of the shutter by pulse control, and the type of results obtained using an exposure time of 2×10^{-6} sec. A reduction of the exposure pulse duration to approximately 10^{-7} sec. could probably be effected if necessary by eliminating the pulse transformer, but a limit of this order is set by the stray capacities associated with the cell and the current rating of the switching valve in the modulator unit.

A reduction of the exposure time would, however, increase the spurious image contribution to the exposure in relation to the normal image contribution and hence the photographic considerations outlined in §3 would be of even greater significance. Clearly any step which can be taken to reduce the contribution of the spurious image is desirable and three possible methods are now proposed. Primarily, an improvement in the 'cut-off' characteristics of the Polaroid elements would contribute towards a reduction of the spurious image component, but whether this could be achieved using the existing type of element material is not known. The use of Nicol prisms as the polarizing elements would also give an improved cut-off, but the difficulty in that case arises in obtaining Nicols of sufficient aperture. A second method for the reduction of the spurious image component is based on the observation that the spectral distribution of the light producing this image is different from the distribution of the light producing the normal image (this observation may not be true for all sources). This difference immediately suggests the introduction of a filter which would reduce the spurious image intensity. (The authors have recently been informed that this technique has been adopted with considerable success by workers of another laboratory.) The third and most involved method depends on the use of a rotating mirror system between the phenomena and the Kerr cell for the purpose of selecting a portion of the emitted light (say over a period of $10 \mu\text{sec.}$), and directing this light towards the Kerr cell, the latter being shielded from direct light from the source. The attendant problems of mirror synchronization and the delay of the selected portion by the required time would require investigation.

Apart from spurious image considerations however, the work has shown that there are at present insufficient data concerning reciprocity-law failure and the Clayden effect for the conditions obtaining in the recording of optical transients by techniques such as described in the paper. Such data are required before the quantitative aspects of recording can be put on a rigorous basis, and further work along these lines would be of value.

ACKNOWLEDGMENTS

In conclusion, the authors wish to express appreciation of the advice given by members of the Ilford and Kodak organizations during the course of the work. Acknowledgment is also due to Professor J. M. Meek and Dr. J. D. Craggs of this Department, who have assisted the authors by discussing the work and this paper on many occasions. Finally, one of us (A. E. J. H.) desires to acknowledge the award of a grant from the Department of Scientific and Industrial Research which enabled the work to be undertaken.

REFERENCES

- ABRAHAM, H., and LEMOINE, J., 1900, *C. R. Acad. Sci., Paris*, **130**, 245.
 BEAMS, J. W., 1926, *J. Opt. Soc. Amer.*, **13**, 597.
 BERG, W. F., 1940, *Proc. Roy. Soc. A*, **174**, 559; 1946, *Photogr. J.*, **86B**, 154.
 BRINKMAN, H., 1937, *Dissertation*, Utrecht.
 DUNNINGTON, F. G., 1931, *Phys. Rev.*, **38**, 1535.
 DUNNINGTON, F. G., and LAWRENCE, E. O., 1930, *Phys. Rev.*, **35**, 396.
 ENGSTROM, R. W., 1947, *J. Opt. Soc. Amer.*, **37**, 420.
 FROOME, K. D., 1948, *J. Sci. Instrum.*, **25**, 371.
 GERMESHAUSEN, K. J., 1948, *Pulse Generators*, Vol. 5 (New York: McGraw-Hill).
 GRIMLEY, T. B., and SAXE, R. F., 1949, *E.R.A. Report L/T211*.
 HÖRMANN, H., 1935, *Z. Phys.*, **97**, 539.
 KINGSBURY, E. F., 1930, *Rev. Sci. Instrum.*, **1**, 22.
 LAPSLEY, A. C., SNODDY, L. B., and BEAMS, J. W., 1948, *J. Appl. Phys.*, **19**, 111.
 MEES, C. E. K., 1945, *Theory of the Photographic Process* (New York: Macmillan).
 POLAROID LABORATORY PRODUCTS (*Bulletin*), Polaroid, Corporation, Cambridge, Mass.
 PRIME, H. A., and SAXE, R. F., 1949, *Proc. Instn. Elect. Engrs.*, Pt. II, **96**, 662.
 RAWCLIFFE, R. D., 1942, *Rev. Sci. Instrum.*, **13**, 413.
 WASHBURN, H. W., 1932, *Phys. Rev.*, **39**, 688.
 WEBB, J. H., 1933, *J. Opt. Soc. Amer.*, **23**, 316; 1935, *Ibid.*, **25**, 4.
 ZAREM, A. M., MARSHALL, F. R., and POOLE, F. L., 1949, *Elect. Engng.*, **68**, 282.

The Optical Emissivity of Titanium and Zirconium

By F. J. BRADSHAW *

Royal Aircraft Establishment, Farnborough

MS. received 11th January 1950

ABSTRACT. A method of measuring the optical emissivity of titanium and zirconium using eddy current heating is described. The specimen consisted of a cylinder of metal drilled with a small black-body hole, and its real and apparent temperatures were measured with an optical pyrometer at a mean wavelength of 0.652μ . The variation of emissivity with temperature was investigated and the precautions taken to avoid errors are described.

As part of a programme of work at the Royal Aircraft Establishment, Farnborough, it was decided to measure the optical emissivities of titanium and zirconium. Existing data consisted only of unpublished work quoted by Van Arkel (1939). Measurements of the variation of vapour pressure with temperature of these metals were also being carried out at Farnborough and for this reason it was felt necessary to check the emissivities.

§ 1. METHOD

Of the various possible ways of measuring emissivity, the direct temperature comparison 'bore hole' method was chosen. That is, a block of material of which the surface emissivity is to be measured has a small cavity of a certain depth made in it and is then heated. If the radiation emerging from the cavity has suffered many internal reflections then it approximates closely to black-body radiation. Measurements of the temperature of the surface and of the cavity with an optical pyrometer lead to the emissivity by

$$\ln \epsilon_\lambda = 1.432/\lambda \left(\frac{1}{T_H} - \frac{1}{T_s} \right),$$

where ϵ_λ is the spectral emissivity, λ the mean effective pyrometer wavelength, T_H the cavity temperature and T_s the surface temperature.

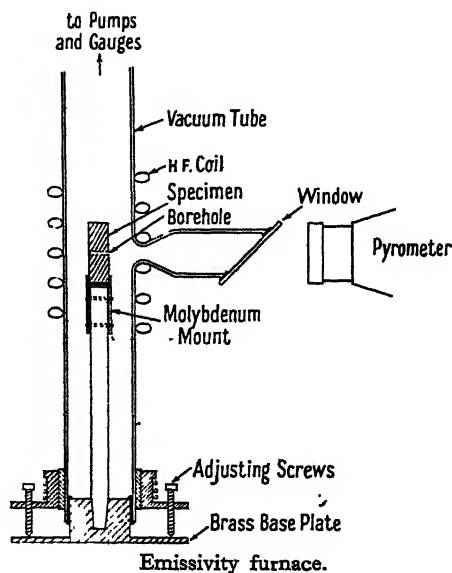
The main difficulties of this method are in producing a uniformly heated specimen and at the same time having it in surroundings the temperatures of which are low enough not to interfere, by reflection, with the specimen's inherent surface brightness. This makes indirect furnace heating unsatisfactory. With electrical resistance heating of the specimen it is still difficult to get a uniform temperature throughout and true black-body radiation from the cavity (Price 1947). However, by using high-frequency eddy current heating these difficulties were overcome. The specimen was a solid cylindrical block of material and was mounted inside an H.F. coil. The circulating currents flowed round the curved surface so that the heat was put in effectively only at the surface. Calculation showed the mean current depth of penetration to be sufficiently small for the cross section temperature gradients to be negligible. Thus the centre temperature was the same as that of the outside, and a bore hole which was drilled down a diameter had a uniform wall temperature throughout its whole length. The only source of non-uniformity was at the ends of the cylinder, where there was nothing but loss by radiation. This non-uniformity was reduced to a negligible amount by the use of a sufficiently long specimen.

* Now at the British Iron and Steel Research Association, London.

§ 2. APPARATUS

The apparatus is shown in the Figure. The metals were obtained in powder form, sintered *in vacuo* (Bickerdike and Sutcliffe 1949) and then suitably worked into fully dense bars. Care was taken to keep the material as pure as possible (for analyses, see § 5). Specimens were turned out of the bars in the form of cylinders 2.0 cm. long and 7.3 mm. in diameter. These dimensions allowed the central portion of the specimen, where the bore hole was drilled, to reach a uniform temperature when heated. The surface of the specimen was polished by normal metallographic methods, magnesia being used in the final stages.

The specimen was held in a molybdenum mount on the top of an alumina rod and the rod was tapered to fit a brass plate which was waxed (Apiezon W 100) into the vacuum tube. The specimen was aligned and centred with the wax still soft by three adjusting screws mounted rigidly to the vacuum tube and bearing on the base plate.



The vacuum tube was made in Pyrex and fitted inside the H.F. coil. A side arm was let into the tube just opposite the intended bore hole position and a plane glass window was waxed on to the end of an angle of 45° to the side arm axis. The pyrometer was then sighted through the window, down the arm at the bore hole. With this arrangement radiation was not reflected back by the concentric walls of the vacuum tube on to the specimen, which would have raised the apparent surface temperature without affecting the bore hole temperature. (These radiation reflection errors could have amounted to as much as 5°C .)

A Metrovic ionization vacuum gauge was mounted directly over the specimen. Pumping was by rotary and oil diffusion pumps with a liquid air trap between them and the vacuum tube.

A G.E.C. 1 kw. portable eddy current heating set, with variable power control, operating at approximately 10 Mc/s. was used, with a heating coil about 5 cm. long, 2.5 cm. diameter and of 5 turns.

A Leeds and Northrup optical pyrometer was used. The transmission curves of the red and neutral filters were measured and the mean effective wavelength was

then calculated to be 0.652μ for the temperature ranges used. The pyrometer was checked at intervals against standard lamps calibrated by the National Physical Laboratory.

§ 3. PROCEDURE

The following preliminary points were investigated using titanium specimens, suitable zirconium being very scarce (in view of the similarity between the two metals, it is unlikely that the conclusions with zirconium would have been different).

To check that the heated specimen was at a uniform temperature, readings were taken at various points on its surface. No change in temperature was detectable at any point on circles perpendicular to the axis of the specimen. A slight temperature gradient parallel to the axis of about $\frac{1}{2}^{\circ}\text{C.}$ per mm. was noted, due probably to differences in coupling between coil and specimen. To avoid any error, measurements of surface temperature were taken on the circle containing the bore hole and close to it. The experimental readings were corrected for pyrometer errors and window absorption errors with the aid of standard lamps. Increased window absorption due to deposit of metallic vapour at higher temperatures was allowed for where necessary.

Buckley (1934) has calculated the theoretical minimum ratio of bore hole length to diameter necessary for the radiation to approximate to that of a black body. His figure was taken as a guide, but confirmatory experiments on holes of various dimensions were also carried out to check it. The final dimensions were: depth 6.0 mm., diameter 0.85 mm. As an added precaution, the inside of the hole was etched with hydrofluoric acid so as to give a pitted roughened surface.

A certain amount of surface thermal etching occurred, and to test its effect, an etched specimen was measured for emissivity, removed, repolished, remounted and again measured. No change was found. Similarly, although all specimens showed grain growth after a run, no change in emissivity with crystal orientation was observed.

§ 4. OBSERVATIONS

Pure titanium and zirconium change from hexagonal to body-centred cubic form at 880°C. and 865°C. respectively (Van Arkel 1939). Work done on zirconium indicated that the change point is spread over a temperature range when impurities such as oxygen are present (Fast 1938). From the analysis of the metal and Fast's work, it was thought that the transition of the zirconium-used would not spread over more than $865\text{--}925^{\circ}\text{C.}$ Similarly it was assumed that the titanium transition was between 880 and 940°C. Figures for emissivity inside these regions were thus rejected. Readings were taken at successively increasing temperatures, the whole run lasting $1\frac{1}{2}$ to 2 hours. Mean pressures were normally about 2×10^{-5} mm. Hg at the start of a run and fell to about 8×10^{-6} mm. at the end. The temperature range was limited at the lower end by the increasing errors due to poor visibility and at the higher end by the power available and the rapid deposition of metallic vapour on the window.

Titanium. From the mean of three runs:

Hexagonal. Temperature range $810\text{--}870^{\circ}\text{C.}$

$\epsilon_{\lambda} = 0.459$ at a mean temperature of 840°C.

No significant variation with temperature was detected.

Body-centred cubic. Temperature range $950\text{--}1,380^{\circ}\text{C.}$

Temp. ($^{\circ}\text{C.}$)	950	1000	1100	1200	1300	1350
ϵ_{λ}	0.484	0.482	0.479	0.476	0.473	0.471

These figures were taken from the mean regression line, that is, the line best fitting the experimental points, of the three runs. Observational errors arose mainly in the readings and in power supply fluctuations. The standard deviation of the absolute value of ϵ_λ was 0.01 (corresponding to about 1.5° c. in temperature).

Zirconium. Work on this was limited by scarcity of material. From the mean of two runs :

Hexagonal. Temperature range 820–840° c.

$\epsilon_\lambda = 0.436$ at a mean temperature of 825° c.

No variation with temperature was detected.

Body-centred cubic. Temperature range 1,020–1,540° c.

$\epsilon_\lambda = 0.426$ at a mean temperature of 1,308° c.

No variation with temperature was detected.

The standard deviation of ϵ_λ was 0.013.

§ 5. SUMMARY AND ANALYSES

The following analyses for likely impurities were made by Messrs. Johnson Matthey and Co. Ltd.

<i>Titanium</i> bar sample :	Iron	0.096%	Magnesium	<0.002%
	Nickel	<0.002%	Oxygen	0.073%
	Silicon	0.056%	Nitrogen	0.19%
	Carbon	0.059%		
<i>Zirconium</i> bar sample :	Iron	0.36%	Magnesium	Nil
	Nickel	0.02%	Oxygen	0.075%
	Silicon	0.13%	Nitrogen	0.17%
	Carbon	0.048%		

The zirconium also contained hafnium, and a recent analysis of material from the same source gave the proportion as 3.0%.

It was known that both metals would adsorb gases during the experiment and one specimen each of titanium and zirconium was analysed for oxygen and nitrogen after a typical run. There was no detectable increase in oxygen or nitrogen in either case. The emissivity of a specimen held at 1,150° c. for 2½ hours (longer than any normal run) showed in the first 25 minutes a barely detectable rise (<0.008) and then remained at a steady value.

These emissivity results are different from those quoted by Van Arkel, who gives:

Titanium at $\lambda = 0.655 \mu$.

Temp. (° c.)	880	980	1080	1180	1280	1380
Emissivity	0.533	0.523	0.502	0.480	0.450	0.419

Zirconium at $\lambda = 0.652 \mu$.

Hexagonal emissivity	0.48
Body-centred emissivity	0.43

The purity of material and conditions of experiment were not given. However, evidence for the correctness of the present results at least for titanium is given by the work of Carpenter and Mair on the vapour pressure of titanium, shortly to be

published. It is found that, using Van Arkel's figures to convert from brightness to true temperature, the latent heat of sublimation deduced from the temperature coefficient of the vapour pressure is inconsistent with that derived from the value of the free energy together with the known difference between the entropy of solid and gas. If the values of emissivity reported here are used to convert from brightness to true temperature the discrepancy disappears.

Michels and Wilford (1949) have recently given a minimum value for titanium emissivity of 0.59 (at $0.665\ \mu$). The figure is certainly high compared with any of the above, but the material was of commercial purity.

ACKNOWLEDGMENTS

I wish to thank my colleagues at the Royal Aircraft Establishment, Farnborough, for their help in this work, and in particular Mr. L. G. Carpenter for his guidance throughout.

This paper is published with the permission of the Chief Scientist, Ministry of Supply, and of the Controller, H.M. Stationery Office.

REFERENCES

- BICKERDIKE, R. L., and SUTCLIFFE, D. A., 1949, *Metallurgia, Manchr.*, **39**, 303.
BUCKLEY, H., 1934, *Phil. Mag.*, **17**, 576.
FAST, J. D., 1938, *Metallwirtschaft*, **17**, 641.
MICHEL, W. C., and WILFORD, S., 1949, *J. Appl. Phys.*, **20**, 1223.
PRICE, D. J., 1947, *Proc. Phys. Soc.*, **59**, 118.
VAN ARKEL, A. E., 1939, *Rein Metalle* (Berlin: Julius Springer), pp. 187 and 202.

Determination of the Elastic and Piezoelectric Coefficients of Monoclinic Crystals, with particular Reference to Ethylene Diamine Tartrate

By R. BECHMANN

Post Office Research Station, Dollis Hill, London

MS. received 14th November 1949

ABSTRACT. The dynamic determination of elastic and piezoelectric coefficients of crystals belonging to the monoclinic sphenoidal class is discussed using the longitudinal mode of vibration for narrow bars, low-frequency longitudinal and face-shear modes for square plates containing the axis of symmetry, coupled modes for square plates perpendicular to the axis of symmetry and the thickness-shear mode for plates containing the axis of symmetry. A new set of values for the elastic and piezoelectric coefficients and their temperature coefficients for ethylene diamine tartrate is given based on this method. Some properties are considered of face-shear vibrating square plates of ethylene diamine tartrate rotated about the axis of symmetry as functions of the orientation, and of Y-cut plates as functions of the breadth-length ratio.

§ 1. INTRODUCTION

THE investigation of the performance of piezoelectric crystals involves the determination of their elastic and piezoelectric coefficients and this becomes more difficult with decreasing symmetry of the crystal. For instance, no example of the determination of the 21 elastic and 18 piezoelectric coefficients of a triclinic crystal (triclinic pedial class) has yet been published. However, the piezoelectric coefficients of several crystals representing the monoclinic sphenoidal class have been determined; e.g. tartaric acid (Tamaru 1905),

cane sugar (Holman 1909), rhamnose (Meyer 1937), lithium sulphate (Spitzer 1938), di-ammonium tartrate (Spitzer 1938), dipotassium tartrate (DKT) (Spitzer 1938, Mason 1946, Jaffe 1948), ethylene diamine tartrate (EDT) (Mason 1947, Jaffe 1948, Bechmann and Lynch 1949). Of these crystals the elastic coefficients are known only for DKT and EDT; they were determined by Mason (1946, 1947), and this was the first determination of these coefficients for monoclinic crystals. The elastic and piezoelectric coefficients have been determined for numerous crystals belonging to classes of higher symmetry.

The method of determining the required coefficients should be the simplest and most reliable for the size and quantity of crystal available. New synthetic materials are usually available at first in small quantities only and this can impose severe limitations on the experimental technique employed. In the following, none of these limitations are assumed, and rough determinations on small crystals are not taken into consideration.

To determine the elastic coefficients of crystals of low symmetry, the simplest method is the dynamic one, in which the constants are deduced from observations of the natural frequencies, dimensions and density of piezoelectrically excited bars and plates made of the material to be tested. The piezoelectric coefficients can be measured statically and can also be determined from dynamic measurements on piezoelectrically excited bars and plates. The piezoelectric coefficients of a number of crystals have been measured by Spitzer (1938) by an improved static method, in which periodically varying stresses were used and for which accuracy within 1% was claimed.

Recently the dynamic methods have been more widely used. Mason (1946) has shown how both the elastic and piezoelectric coefficients can be determined dynamically, the elastic coefficients from measurements of resonance frequencies, and the piezoelectric coefficients by measurements of frequencies of resonance and anti-resonance. All the measurements were to be made on bars.

This paper describes a dynamic method in which the elastic and piezoelectric shear coefficients are deduced from the properties of square plates. This has the advantage of determining these constants more accurately and with greater reliability. The formulae for determination of the piezoelectric coefficients are given in terms of the dynamic capacitance in the equivalent circuit of the resonator, because the magnitude of this circuit element can be measured with satisfactory accuracy in several ways.

The following considerations are quite general for the monoclinic spenoidal class. In accordance with the conventions for piezoelectric crystals the binary axis is taken as the b axis and as Y axis for the rectangular coordinate system used for the physical properties. The schemes of elastic, s_{ik} , and piezoelectric, d_{ik} , coefficients and permittivities, ϵ_{ik} , referred to this coordinate system, the Y axis being the axis of symmetry, are:

$$\begin{array}{cccccccccccc}
 s_{11} & s_{12} & s_{13} & 0 & s_{15} & 0 & 0 & 0 & 0 & d_{14} & 0 & d_{16} \\
 & s_{22} & s_{23} & 0 & s_{25} & 0 & d_{21} & d_{22} & d_{23} & 0 & d_{25} & 0 \\
 & & s_{33} & 0 & s_{35} & 0 & 0 & 0 & 0 & d_{34} & 0 & d_{36} \\
 & & & s_{44} & 0 & s_{46} & & & & & & \\
 & & & & s_{55} & 0 & & & & & & \\
 & & & & & s_{66} & & & & & & \\
 \epsilon_{11} & \epsilon_{22} & \epsilon_{33} & 0 & \epsilon_{13} & 0 & & & & & & \dots\dots(1)
 \end{array}$$

The scheme of coefficients of linear expansion a_{ik} corresponds to that of the permittivities ϵ_{ik} .

For the determination of all elastic and piezoelectric coefficients using the dynamic method, the following modes of vibration are used and the formulae for these modes are given in detail: longitudinal mode for narrow bars, low-frequency longitudinal and face-shear modes for square plates containing the axis of symmetry (Y axis), coupled modes for square plates perpendicular to the axis of symmetry, and thickness-shear mode for plates containing the axis of symmetry. For definitions of these modes see Cady (1946).

For crystals with higher symmetry, in general similar considerations hold good, but there are fewer coefficients, and the formulae are simpler on this account.

§2. LONGITUDINAL MODE FOR NARROW BARS

Some of the elastic and piezoelectric coefficients can be determined from measurements of the longitudinal mode of vibration of small bars. Provided that the surfaces perpendicular to the electric field, which is in the direction of the thickness, are fully covered with very thin conducting layers as electrodes, the other surfaces being free from conducting layers, the frequency of series resonance for a bar parallel to the Y' axis is given by

$$\nu = \frac{1}{2l} \left(\frac{1}{\rho s_{22}'} \right)^{1/2} \left(1 + \frac{2\pi}{\epsilon_{22}'} \cdot \frac{d_{22}'^2}{s_2'} + \frac{2\pi}{\epsilon_{11}'} \cdot \frac{d_{12}'^2}{s_{22}'} \right), \quad \dots\dots(2)$$

where l is the length of the bar and ρ its density, s_{22}' is the elastic coefficient in the direction of the length of the bar, d_{12}' , d_{22}' are piezoelectric coefficients, ϵ_{22}' , ϵ_{11}' are permittivities. The dashed symbols refer to any chosen direction of the orientation of the axis of the bar.

The frequency equation (2) contains two correction terms; first, if d_{22}' is not zero, space charges distributed along the length of the bar appear, and thus the elastic coefficient is changed; second, if the piezoelectric coefficient d_{12}' in the direction of the breadth of the bar is not zero, there is another term due to the boundary conditions caused by the surface charges on these faces.

The piezoelectric coefficient d_{32}' is given in terms of the frequency ν and the dynamic capacitance C_k of the equivalent electric circuit (in E.S.U.)

$$d_{32}' = \left(\frac{\pi^2}{32\rho} \cdot \frac{C_k}{\nu^2} \cdot \frac{t}{bl^3} \right)^{1/2}, \quad \dots\dots(3)$$

where b and t are the breadth and thickness of the bar.

To define the orientation of the bar in the coordinate system given, a polar coordinate system is used as shown in Figure 1, the Y axis (binary axis) being polar axis. A plane containing the Y axis, rotated from the positive Z axis to the positive X axis, is determined by the angle ϕ . For a bar with its length in this plane, the axis of the bar is defined by the angle θ , measured from the positive Y axis. The angle ψ describes a rotation of the bar about its own axis. It gives the direction of the piezoelectric excitation. $\psi=0$ means that the direction of the electric field falls in the plane considered containing the Y axis. For $\psi=90^\circ$ the electric field is perpendicular to this plane. The angle ψ therefore indicates the particular surface intended to carry the electrodes. In the above formulae the direction of the electric field corresponds to the thickness of the bar.

The well-known general expressions for s_{ik}' and d_{ik}' as functions of the direction cosines are simplified for the particular case of monoclinic sphenoidal class with the scheme (1) of s_{ik} and d_{ik} . Using the above-mentioned polar coordinate system with the Y axis as the polar axis, the following equations for the coefficients s_{22}' , d_{32}' , d_{22}' and the permittivities ϵ_{22}' , ϵ_{11}' are obtained:

$$\begin{aligned} s_{22}' = & [s_{33} \cos^4 \phi + s_{11} \sin^4 \phi + (s_{55} + 2s_{13}) \sin^2 \phi \cos^2 \phi \\ & + 2s_{15} \sin^3 \phi \cos \phi + 2s_{35} \sin \phi \cos^3 \phi] \sin^4 \theta \\ & + [(s_{44} + 2s_{23}) \cos^2 \phi + (s_{66} + 2s_{12}) \sin^2 \phi \\ & + 2(s_{25} + s_{46}) \sin \phi \cos \phi] \sin^2 \theta \cos^2 \theta + s_{22} \cos^4 \theta \end{aligned} \quad \dots\dots (4)$$

$$\begin{aligned} d_{32}' = & \cos \psi [\{d_{23} \cos^2 \phi + d_{21} \sin^2 \phi + d_{25} \sin \phi \cos \phi\} \sin^3 \theta \\ & + \{(d_{22} - d_{34}) \cos^2 \phi + (d_{22} - d_{16}) \sin^2 \phi \\ & - (d_{14} + d_{36}) \sin \phi \cos \phi\} \sin \theta \cos^2 \theta] \\ & - \sin \psi [d_{14} \cos^2 \phi - d_{36} \sin^2 \phi + (d_{16} - d_{34}) \sin \phi \cos \phi] \sin \theta \cos \theta \end{aligned} \quad \dots\dots (5)$$

$$\begin{aligned} d_{22}' = & [(d_{23} + d_{34}) \cos^2 \phi + (d_{21} + d_{16}) \sin^2 \phi + (d_{14} + d_{25} + d_{36}) \sin \phi \cos \phi] \\ & \times \sin^2 \theta \cos \theta + d_{22} \cos^3 \theta \end{aligned} \quad \dots\dots (5a)$$

$$\begin{aligned} \epsilon_{11}' = & \cos^2 \psi [\epsilon_{11} \cos^2 \phi + \epsilon_{33} \sin^2 \phi - 2\epsilon_{13} \sin \phi \cos \phi] \\ & + \sin^2 \psi [\{\epsilon_{33} \cos^2 \phi + \epsilon_{11} \sin^2 \phi + 2\epsilon_{13} \sin \phi \cos \phi\} \cos^2 \theta + \epsilon_{22} \sin^2 \theta] \\ & - \sin \psi \cos \psi [(\epsilon_{11} - \epsilon_{33}) \sin 2\phi + 2\epsilon_{13} \cos 2\phi] \cos \theta \end{aligned} \quad \dots\dots (6)$$

$$\epsilon_{22}' = [\epsilon_{33} \cos^2 \phi + \epsilon_{11} \sin^2 \phi + 2\epsilon_{13} \sin \phi \cos \phi] \sin^2 \theta + \epsilon_{22} \cos^2 \theta. \quad \dots\dots (6a)$$

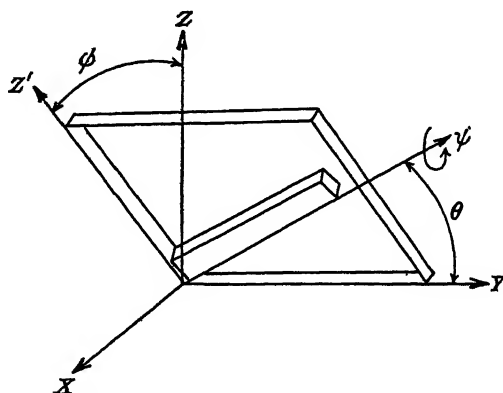


Figure 1. Diagram illustrating the coordinates used to define the orientation of bars.

The formulae (4), (5a) and (6a) for s_{22}' , d_{22}' and ϵ_{22}' are independent of ψ . The coefficient d_{12}' appearing in equation (2) is obtained from d_{32}' (equation (5)) by writing $\pi/2 + \psi$ for ψ .

For bars in various coordinate and inclined planes the formulae simplify to

(i) bars in the YZ plane: $\phi = 0$

$$s_{22}' = s_{22} \cos^4 \theta + s_{33} \sin^4 \theta + (s_{44} + 2s_{23}) \sin^2 \theta \cos^2 \theta,$$

$$\psi = 0: d_{32}' = d_{23} \sin^3 \theta + (d_{22} - d_{34}) \sin \theta \cos^2 \theta,$$

$$\psi = 90^\circ: d_{32}' = -d_{14} \sin \theta \cos \theta;$$

(ii) bars in the XY plane: $\phi = 90^\circ$

$$s_{22}' = s_{22} \cos^4 \theta + s_{11} \sin^4 \theta + (s_{66} + 2s_{12}) \sin^2 \theta \cos^2 \theta,$$

$$\psi = 0: d_{32}' = d_{21} \sin^3 \theta + (d_{22} - d_{16}) \sin \theta \cos^2 \theta,$$

$$\psi = 90^\circ: d_{32}' = d_{36} \sin \theta \cos \theta;$$

(iii) bars in the XZ plane: $\theta = 90^\circ$

$$s_{22}' = s_{33} \cos^4 \phi + s_{11} \sin^4 \phi + (s_{55} + 2s_{13}) \sin^2 \phi \cos^2 \phi \\ + 2s_{15} \sin^3 \phi \cos \phi + 2s_{35} \sin \phi \cos^3 \phi,$$

$$\psi = 0: d_{32}' = d_{23} \cos^2 \phi + d_{21} \sin^2 \phi + d_{15} \sin \phi \cos \phi,$$

$$\psi = 90^\circ: d_{32}' = 0;$$

(iv) bars in the plane: $\phi = 45^\circ$

$$s_{22}' = \frac{1}{4}[s_{11} + s_{33} + s_{55} + 2s_{13} + 2s_{15} + 2s_{35}] \sin^4 \theta \\ + \frac{1}{2}[s_{44} + 2s_{23} + s_{66} + 2s_{12} + 2(s_{25} + s_{46})] \sin^2 \theta \cos^2 \theta \\ + s_{22} \cos^4 \theta,$$

$$\psi = 0: d_{32}' = \frac{1}{2}[(d_{21} + d_{23} + d_{25}) \sin^3 \theta + \{(d_{22} - d_{34}) + (d_{22} - d_{16}) \\ - (d_{14} + d_{36})\} \sin \theta \cos^2 \theta],$$

$$\psi = 90^\circ: d_{32}' = -\frac{1}{2}[(d_{14} - d_{36}) + (d_{16} - d_{34})] \sin \theta \cos \theta.$$

Using the measurements of frequency, capacitance and dimensions of an adequate number * of suitably orientated bars in the planes considered above and solving the appropriate equations it is possible to determine: 5 elastic coefficients: s_{11} , s_{22} , s_{33} , s_{15} , s_{35} ; 4 combinations of elastic coefficients: $s_{44} + 2s_{23}$, $s_{55} + 2s_{13}$, $s_{66} + 2s_{12}$, $s_{25} + s_{46}$; 5 piezoelectric coefficients: d_{21} , d_{23} , d_{25} , d_{14} , d_{36} ; 3 combinations of piezoelectric coefficients: $d_{22} - d_{16}$, $d_{22} - d_{34}$, $d_{16} - d_{34}$.

§ 3. LOW-FREQUENCY LONGITUDINAL AND FACE-SHEAR MODES OF SQUARE PLATES CONTAINING THE AXIS OF SYMMETRY (Y AXIS)

Generally, square and rectangular plates of any orientation can be excited by an electric field normal to the plates in modes dependent on the contour, and these are modified by the coupling of longitudinal and shear modes (Bechmann 1941, 1942). In particular, square plates having their major planes and one edge parallel to the axis of symmetry can only be excited in a pure shear mode given by the elastic coefficient s_{44}' because s_{24}' and s_{34}' , which cause the coupling between longitudinal and shear modes, both vanish. If one edge is inclined at 45° to the

* The signs of the piezoelectric coefficients must be regarded as an additional unknown quantity, see § 6.

axis of symmetry, then $s_{22}' = s_{33}'$ and $s_{24}' = s_{34}'$ and the plate can be excited in a longitudinal mode *, the resultant elastic coefficient being $s_{22}' - s_{23}'$. For two square plates in the same plane, one inclined 45° to the axis of symmetry and the other parallel to this axis (0°), there is the relation

$$(s_{22}' - s_{23}')_{45^\circ} = \frac{1}{2}(s_{44}')_{0^\circ} \quad \text{and} \quad (d_{12}')_{45^\circ} = -(d_{13}')_{45^\circ} = \frac{1}{2}(d_{14}')_{0^\circ}.$$

Using the above-mentioned polar coordinate system the plate parallel to the Y axis is defined by the angle Θ between the normal of the plate and the X axis. The angle Θ is measured from the positive direction of the X axis towards the negative direction of the Z axis. The elastic and piezoelectric coefficients s_{44}' and d_{14}' as functions of the angle Θ are

$$\left. \begin{aligned} s_{44}' &= s_{44} \cos^2 \Theta + s_{66} \sin^2 \Theta + 2s_{46} \sin \Theta \cos \Theta, \\ d_{14}' &= d_{14} \cos^2 \Theta - d_{36} \sin^2 \Theta + (d_{16} - d_{34}) \sin \Theta \cos \Theta. \end{aligned} \right\} \dots\dots (7)$$

The expression for the series resonance frequency for the longitudinal and shear modes of square plates can be written in the same form:

$$\nu = \frac{a\sqrt{2}}{2l} \sqrt{\left(\frac{1}{\rho s_{44}'}\right)}, \quad \dots\dots (8)$$

where l is the length of the square. The factor a is equal to 1 for the longitudinally vibrating plates inclined at 45° to the axis of symmetry; formula (8) is an exact solution for this case. The factor a for the shear mode in equation (8) was first determined empirically from quartz (Hight and Willard 1937, Bechmann 1942). For an isotropic plate Ekstein (1944) gives the value $a = a_0 = 0.9132$; Mähly and Trösch (1947) give $a_0 = x\sqrt{2}/\pi$ where x is root of equation

$$\tan x + x = 0. \quad \dots\dots (9)$$

$x = 2.0288$ is the first root and therefore $a_0 = 0.9132$. An improved theory of shear vibration of square plates by Mähly and Trösch (1947) taking into consideration longitudinal effects gives a corrected value $a = a_0(1 - A\sqrt{\sigma})$ and A becomes

$$A = \frac{1}{2x} \left(\frac{x^2 - 2}{x^2 + 2} \right)^{3/2} = 0.0501_2,$$

where for an isotropic plate ($s_{22} = s_{33}'$), $\sigma = s_{22}/s_{44}$. In the case considered where $s_{22} \neq s_{33}'$ an average value between σ formed with s_{22} and s_{33}' can be used. This value for a is only an approximation, but it has been shown experimentally to be reliable. If both kinds of plates are used and N_{0° means the frequency constant (i.e. the product of frequency and length) for the plate with one edge parallel to the Y axis, and N_{45° the frequency constant for the plate at the same angle Θ but inclined at 45° to the Y axis, then

$$\left(\frac{N_{0^\circ}}{N_{45^\circ}} \right)_\Theta = a. \quad \dots\dots (10)$$

* This longitudinal mode agrees with one of the longitudinal modes (mode 1) derived by Ekstein (1944) whose equation (48) for a Z-cut quartz plate, or an isotropic plate, is $\rho\omega^2 = k^2(\gamma_{11} - \gamma_{12})$ with

$$\gamma_{11} = \frac{s_{11}}{s_{11}^2 - s_{12}^2}, \quad \gamma_{12} = \frac{-s_{12}}{s_{11}^2 - s_{12}^2}, \quad \text{that is} \quad \gamma_{11} - \gamma_{12} = \frac{1}{s_{11} - s_{12}} = \frac{1}{\frac{1}{2}s_{66}}.$$

Another mode can also be excited in a plate rotated about 45° to the axis of symmetry which corresponds with Ekstein's longitudinal mode 2.

According to the above expressions, a is slightly dependent on the direction of the particular orientation Θ and also on the material.

For the longitudinally vibrating square plate inclined at 45° to the axis of symmetry the piezoelectric coefficient d_{14}' expressed in terms of series resonance frequency ν and the capacitance C_k is given by*

$$d_{14}' = \left(\frac{\pi^4}{128\rho} \cdot \frac{C_k}{\nu^2} \cdot \frac{t}{l^4} \right)^{1/2}, \quad \dots\dots(11)$$

where t is the thickness of the plate. In the case of the shear vibrating square plate parallel to the axis the expression for d_{14}' becomes

$$d_{14}' = \left(\frac{(x^2 + 2)x^2}{4\pi^2\rho} \cdot \frac{C_k}{\nu^2} \cdot \frac{t}{l^4} \right)^{1/2}; \quad \dots\dots(12)$$

x means the root of equation (9).

Summarizing, from measurements of these plates, the following coefficients can be obtained: 3 elastic coefficients s_{44} , s_{66} , s_{46} , 2 piezoelectric coefficients d_{14} , d_{36} , and the combination of the piezoelectric coefficients $d_{16} - d_{34}$.

§ 4. SQUARE PLATES WITH COUPLED MODES PERPENDICULAR TO THE AXIS OF SYMMETRY (Y AXIS)

For the determination of s_{55} and s_{13} which are known from § 2 as a combination $s_{55} + 2s_{13}$ it is necessary to use bars or plates vibrating in a coupled mode. A simple case is a square plate perpendicular to the axis of symmetry (Y axis), having edges parallel to the X and Z axes. An approximate frequency equation recently given for three of the main modes of a square plate (Bechmann 1942) is

$$\begin{vmatrix} s_{11} - s & s_{13} & s_{15} \\ s_{13} & s_{33} - s & s_{35} \\ s_{15} & s_{35} & s_{55} - 2a^2s \end{vmatrix} = 0, \quad \dots\dots(13)$$

where $s = 1/4\rho\nu^2l^2$, l is the edge-length of the square and a the factor already mentioned in § 3. The only unknown elastic coefficients in this equation are s_{55} and s_{13} . From measurements of the three frequencies the elastic coefficients s_i ($i = 1, 2, 3$) can be found, and these should agree with the roots of the equation (13). There are then two equations containing s_{55} and s_{13} and the solution can be obtained as a cubic in either one. On the other hand the expression $s_{11} + s_{33} + s_{55}/2a^2$ is an invariant of the cubic equation (13), and since this equals the sum of $s_1 + s_2 + s_3$ the elastic coefficient s_{55} can be obtained. It must be checked, however, that s_1 , s_2 and s_3 equal the roots of the frequency equation (13). Therefore, to be quite sure, both forms of determination should be used. The errors in s_{55} and s_{13} include any error of the value a .

Hence all the elastic coefficients can be determined.

§ 5. THICKNESS SHEAR MODE FOR PLATES CONTAINING THE AXIS OF SYMMETRY

The piezoelectric coefficients d_{18} , d_{22} and d_{34} are combined with each other in all the above formulae. To solve these combinations, at least one of them must be determined. This can be done by means of longitudinally vibrating bars parallel to the Y axis, excited by a field along their length; the piezoelectric

* Details of derivation of the equations (11) and (12) will be given in a later paper.

coefficient is then d_{22} . The value for the capacitance becomes very small for narrow bars. A more convenient method uses plates mentioned before (§3), rotated about the Y axis and excited in a thickness mode. Only one mode, the thickness-shear mode, is excitable. The formulae for frequency and capacitance are expressed in terms of elastic and piezoelectric moduli, c_{ik} and e_{ih} , instead of s_{ik} and d_{ik} .

For the thickness-shear mode of the plate *

$$\nu = \frac{1}{2t} \left(\frac{c_{66}'}{\rho} \right)^{1/2} \left(1 + \frac{2\pi e_{16}'^2}{\epsilon_{11}' c_{66}'} - \frac{16 e_{16}'^2}{\pi \epsilon_{11}' c_{66}'} \right). \quad \dots\dots(14)$$

The piezoelectric modulus e_{16}' , given in terms of series resonance frequency ν and capacitance C_k is:

$$e_{16}' = \left(\frac{\pi^2 \rho}{2} \cdot C_k \nu^2 \cdot \frac{t^3}{A} \right)^{1/2}, \quad \dots\dots(15)$$

where A is the area of the plate covered with a thin conducting layer, and t the thickness. In the case of the plates referred to, rotated about the Y axis at the angle Θ ,

$$\left. \begin{aligned} c_{66}' &= c_{66} \cos^2 \Theta + c_{44} \sin^2 \Theta - 2c_{46} \sin \Theta \cos \Theta, \\ e_{16}' &= e_{16} \cos^2 \Theta + e_{34} \sin^2 \Theta - (e_{14} + e_{36}) \sin \Theta \cos \Theta, \\ \epsilon_{11}' &= \epsilon_{33} \cos^2 \Theta + \epsilon_{11} \sin^2 \Theta - 2\epsilon_{13} \sin \Theta \cos \Theta. \end{aligned} \right\} \quad \dots\dots(16)$$

Between the elastic moduli c_{44} , c_{66} , c_{46} and the elastic coefficients s_{44} , s_{66} , s_{46} , and the piezoelectric moduli e_{14} , e_{16} , e_{34} , e_{36} and the piezoelectric coefficients d_{14} , d_{16} , d_{34} , d_{36} , there are the relations

$$\left. \begin{aligned} c_{44} &= s_{66} \Delta \\ c_{66} &= s_{44} \Delta \\ c_{46} &= -s_{46} \Delta \\ \frac{1}{\Delta} &= s_{44} s_{66} - s_{46}^2 \end{aligned} \right\} \begin{aligned} s_{44} &= c_{66} \Delta' \\ s_{66} &= c_{44} \Delta' \\ s_{46} &= -c_{46} \Delta' \\ \frac{1}{\Delta'} &= c_{44} c_{66} - c_{46}^2 \end{aligned} \quad \dots\dots(17)$$

and

$$\begin{aligned} e_{14} + e_{28} &= d_{14} c_{44} + d_{36} c_{66} + (d_{16} + d_{34}) c_{46} & d_{16} + d_{34} &= e_{34} s_{44} + e_{16} s_{66} + (e_{14} + e_{36}) s_{46}. \\ & & & \dots\dots(18) \end{aligned}$$

These results combined with those already obtained give the piezoelectric coefficients d_{16} , d_{22} , d_{34} . Hence all the piezoelectric coefficients can be determined. There can also be found 3 elastic moduli: c_{44} , c_{66} , c_{46} , and 4 piezoelectric moduli: e_{14} , e_{16} , e_{34} , e_{36} .

* Cady (1946, p. 315) quotes an error in the author's former publication (1940) where the additional term for the elastic modulus produced by the piezoelectric effect in formulae (30) and (32 a) is given as $c + 8\pi e^2/\epsilon$. As Cady points out, this term should read $c + 4\pi e^2/\epsilon$. The electric field is not independent and cannot therefore be regarded as an independent variable. This introduces the additional factor 2.

§ 6. MEASUREMENTS AND ACCURACY OF THE DYNAMICALLY DETERMINED COEFFICIENTS

For an exact determination of the coefficients, various influences must be taken into account. The first question is whether the condition in which the bars and plates were measured agrees with the condition on which the formulae were based. For example, bars must be small enough for errors due to the influence of the breadth-length or thickness-length ratio to be unimportant. It must be checked carefully that no disturbing frequencies are present which might influence the results. In particular, when bars having certain dimensional ratios are vibrating longitudinally, flexural modes can be excited by coupling. For example, in a bar of any material of which the width is about 0.21 times the length (see § 7(ii)), the frequency of the second flexural mode coincides with that of the longitudinal mode. If the coupling coefficient between the longitudinal and shear modes is not zero, the properties of the coupled vibration differ from those of the longitudinal vibration, and ratios near this critical value should therefore be avoided. To avoid disturbances due to coupling with other modes, it may occasionally be necessary to alter the dimensions of the specimens.

Using bars and plates belonging to any group considered above more than the minimum number of differently orientated specimens should be measured, so that average values can be determined and an estimate made of the accuracy of the resulting values. Some of the coefficients can be determined independently from different modes of vibration. The values for piezoelectric coefficients determined from measurement on crystals vibrating in different modes show a slight systematic discrepancy which seems due to the fundamental equations of vibration on which the formulae were based. The maximum variation, assuming perfect measurements, is in the order of 1 to 3%. It is thought that the correct value is most likely to correspond with results obtained from longitudinally vibrating square plates. This will be discussed in a later paper already mentioned.

The dynamic method for the determination of piezoelectric coefficients, which only gives the square of these coefficients and in general the square of a combination of such values, results in an ambiguity in the sign which must be taken into account. The sign of one of the coefficients d_{ik} must be chosen according to an arbitrary definition in terms of static effect.

In all formulae for the determination of s_{ik} and d_{ik} , the expressions ρs_{ik} and $\sqrt{\rho} d_{ik}$ occur, so that all values have systematic errors equal to those of ρ and $\sqrt{\rho}$ respectively. The following errors and uncertainties should be taken into consideration: specimens prepared from unsatisfactory material, errors in orientation, inaccuracy of dimensions, inaccuracy of the shapes of bars and plates, influence of silver layer and of mounting, and perhaps influence of the atmospheric pressure and humidity. These inaccuracies relate to values of s_{ik}' and d_{ik}' . The inaccuracy may become numerically greater for values s_{ik} and d_{ik} obtained from combinations of s_{ik}' and d_{ik}' .

The temperature coefficients $T(s_{ik}')$ of the elastic coefficients can be determined from the measured temperature coefficients $T(\nu)$ of frequency by the formula

$$T(s_{ik}') = \frac{1}{s_{ik}'} \frac{\partial s_{ik}'}{\partial T} = -2T(\nu) + \beta_1. \quad \dots\dots(19)$$

Similarly the temperature coefficients $T(d_{ik}')$ of the piezoelectric coefficients can be obtained from measured temperature coefficients $T(C_k)$ and $T(\nu)$ of the capacitance and frequency respectively from the equation

$$T(d_{ik}') = \frac{1}{d_{ik}'} \frac{\partial d_{ik}'}{\partial T} = \frac{1}{2} T(C_k) - T(\nu) + \beta_2, \quad \dots (20)$$

where β_1 and β_2 are functions of the coefficients of linear expansion which depend on the mode of vibration and the orientation of the bars or plates. The temperature coefficients $T(s_{ik})$ and $T(d_{ik}')$ can be found from the $T(s_{ik}')$ and $T(d_{ik}')$ by solving the appropriate equations in the same way as s_{ik} and d_{ik} were obtained from the observed values s_{ik}' and d_{ik}' .

§ 7. RESULTS ON ETHYLENE DIAMINE TARTRATE

A new determination of elastic and piezoelectric coefficients and their temperature coefficients was made on anhydrous ethylene diamine tartrate. The values for the capacitance of the specimens were obtained from measurements of the effective capacitance near resonance. This method has been described by Lynch (1950).^{*} Table 1 contains the values for the adiabatic elastic coefficients.

Table 1. Elastic Coefficients for EDT in 10^{-13} cm² dyne⁻¹

s_{ik}	Mason	New values	s_{ik}	Mason	New values	s_{ik}	Mason	New values
s_{11}	38.8	33.4	s_{23}	-19	-18	s_{15}	-5	-17
s_{22}	37.6	36.5	s_{44}	187	191.8	s_{25}	1	15
s_{33}	98	100.2	s_{55}	171.5	116.8	s_{35}	-25	-26.5
s_{12}	3	-3	s_{66}	173	191.4	s_{46}	-2	3.8
s_{13}	-52	-30						

The error in the values of the main coefficients is of the order of 0.5 to 2% and that of the other coefficients does not exceed 5%. Table 1 also contains values given previously by Mason (1947) which agree substantially, but there are differences for s_{55} , s_{12} , s_{13} , s_{15} and s_{25} . Table 2 shows the values for the temperature

Table 2. Temperature Coefficients of the Elastic Coefficients for EDT in $10^{-6}/^{\circ}\text{C}$.

$T(s_{ik})$		$T(s_{ik})$		$T(s_{ik})$		$T(s_{ik})$	
$T(s_{11})$	290	$T(s_{13})$	-2280	$T(s_{55})$	-280	$T(s_{25})$	-730
$T(s_{22})$	640	$T(s_{33})$	-55	$T(s_{66})$	320	$T(s_{35})$	2220
$T(s_{33})$	-95	$T(s_{44})$	260	$T(s_{15})$	-85	$T(s_{46})$	-28600
$T(s_{12})$	150						

Table 3. Piezoelectric Coefficients for EDT in 10^{-8} E.S.U. dyne⁻¹

d_{ik}	Mason	Jaffe	New values	d_{ik}	Mason	Jaffe	New values
d_{14}	23.5		-30.4	d_{23}	31	-36.9 ± 3	-34.0
d_{15}	37		-36.6	d_{25}	14	-55.2 ± 6	-53.8
d_{21}	34	$+33.9 \pm 3$	+30.3	d_{34}	37.5		-50.9
d_{22}	-20	$+5.4 \pm 1.5$	+6.6	d_{35}	50		-55.1

^{*} The measurements of capacitance of various specimens were made by A. C. Lynch.

coefficients $T(s_{ik})$ of the elastic coefficients, which are accurate to within about 10%. The temperature coefficient for s_{46} has an unusually high negative value. Table 3 contains values of the piezoelectric coefficients, which are accurate to within about 3%. It also includes values first given by Mason (1947) and by Jaffe (1948); Jaffe's measurements were made by an electrostatic method. The new values quoted for 20° c. are practically identical with those recently published (Bechmann and Lynch 1949), but the values for d_{14} and d_{36} are improved. Some of the values of the temperature coefficients for the piezoelectric coefficients $T(d_{ik})$ and the values for the combinations $T(d_{22}-d_{16})$, $T(d_{16}-d_{34})$, $T(d_{22}-d_{34})$ are given in Table 4. The error in these values is less than 10%.

Table 4. Temperature Coefficients of Piezoelectric Coefficients
for EDT in $10^{-4}/^{\circ}\text{C}$.

$T(d_{ik})$	$T(d_{ik})$	$T(d_{ik})$
$T(d_{14}) - 7.3$	$T(d_{23}) - 47$	$T(d_{22}-d_{34}) - 25$
	$T(d_{25}) - 10$	$T(d_{16}-d_{34}) - 48$
$T(d_{21}) - 18.5$	$T(d_{36}) - 15.2$	$T(d_{22}-d_{16}) - 21$

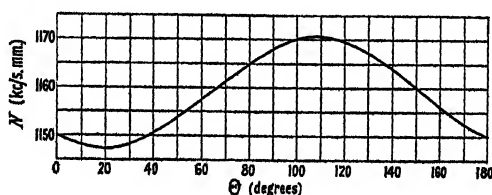


Figure 2. Variation of frequency constant with orientation for shear vibrating square plates of ethylene diamine tartrate rotated about the Y axis.

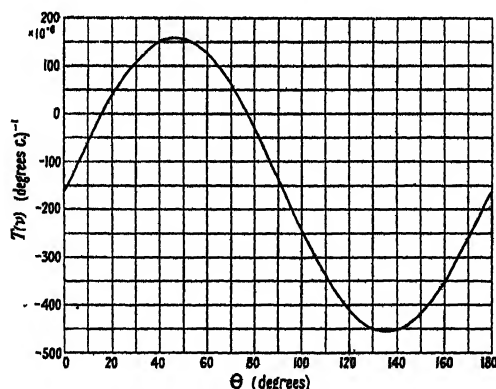


Figure 3. Variation of temperature coefficient of frequency for shear vibrating square plates of ethylene diamine tartrate rotated about the Y axis.

The values are based on the value for the density (see *Analytical Chemist* 1948) $\rho = 1.538 \text{ gm. cm}^{-3}$ and the coefficients of linear expansion according to Mason's values: $a_{11} = 0$, $a_{22} = 20.3$, $a_{33} = 80$, $a_{13} = -32 \times 10^{-6}/^{\circ}\text{C}$. The values for the permittivities are required only in the correction terms of equations (2) and (14), and the values $\epsilon_{11} = 5$, $\epsilon_{22} = 8.2$, $\epsilon_{33} = 6$, $\epsilon_{13} = 0.65$ were used. The temperature coefficients are average values in the range of about 20 to 50° c.

Some interesting special cases may be derived from these values:

(i) *Square plates rotated about the Y axis, vibrating in contour shear mode.*

The frequency constant $N = \nu l$ (equation (8)) is shown as a function of the angle Θ in Figure 2; the temperature coefficient of frequency is given in Figure 3. At the angles $\Theta = 17^\circ$ and 77° the temperature coefficient of frequency is zero (Bechmann 1949). Although the temperature coefficients $T(s_{44})$ and $T(s_{66})$ have the same sign, the zero points are caused by the high negative value of $T(s_{46})$ which balances the values $T(s_{44})$ and $T(s_{66})$ at these angles. The coefficient of inductance $L_0 = L/t$ derived from equation (12), where $L = 1/4\pi^2\nu^2C_k$, is given in Figure 4,

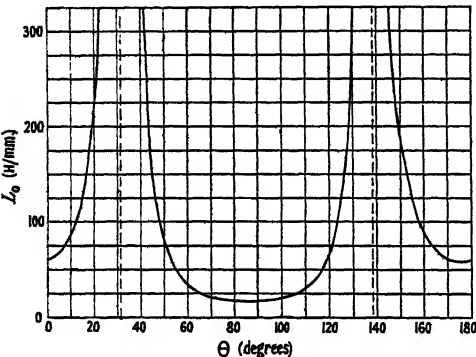


Figure 4. Variation of inductance coefficient for shear vibrating square plates of ethylene diamine tartrate rotated about the Y axis.

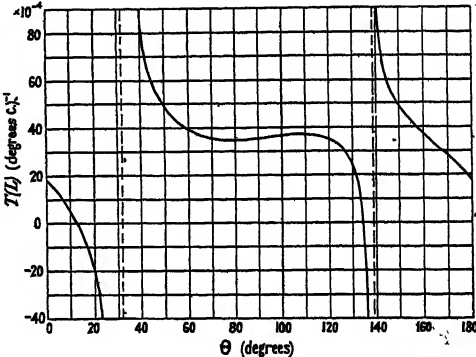


Figure 5. Variation of temperature coefficient of inductance for shear vibrating square plates of ethylene diamine tartrate rotated about the Y axis.

as function of the angle Θ . Finally, Figure 5 shows the temperature coefficient of L which is zero for $\Theta = 13^\circ$ and 136° . All these curves have been calculated from the values given above. The factor a in equation (8) has the observed values for $\Theta = 0^\circ$: $a = 0.882$; $\Theta = 77^\circ$: $a = 0.898$; $\Theta = 90^\circ$: $a = 0.894$.

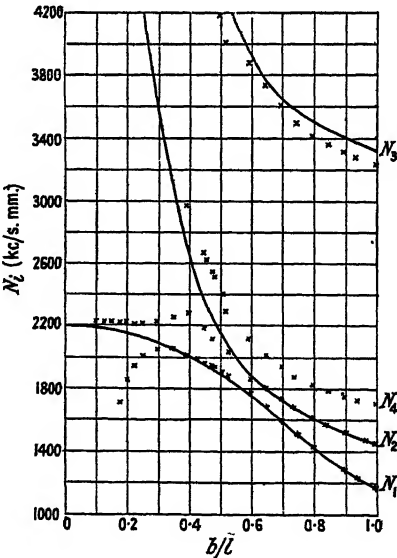


Figure 6. Variation of frequency constants with breadth-length ratio for Y cut plates of ethylene diamine tartrate.

(ii). *Plate perpendicular to the Y axis.*

The calculated and measured frequency constants N_i ($i=1, 2, 3$) of the three vibrations of a plate perpendicular to the Y axis as mentioned before in §4, are given in Figure 6 as a function of the breadth-length ratio, the reduced dimension being perpendicular to the X axis. This calculation was made in the same way as recently for quartz plates (Bechmann 1942). The curves are plotted from calculated results and crosses refer to observed values. For very small values of b/l , N_1 alone remains finite, as this represents the longitudinal vibration of a narrow bar. As the frequency for the longitudinal mode of a small bar is approximately

$$\nu = \frac{1}{2l} \sqrt{\left(\frac{1}{\rho s}\right)}$$

and that for the second flexural mode

$$\nu = \frac{\kappa^2}{4\pi\sqrt{3}} \frac{b}{l^2} \sqrt{\left(\frac{1}{\rho s}\right)},$$

both frequencies coincide when $b/l=0.21$, as $\kappa=7.20$ (Mason 1942) for this ratio. When the longitudinal mode is coupled through the shear mode to the second flexural mode, the coupling effect shown in Figure 6 at about this ratio occurs. This effect is already known in quartz bars. However, it is not included in the calculation for N as shown in Figure 6. N_3 represents the coupled mode, given by the breadth of the plate, and N_4 is unidentified, but this vibration may be related to mode 2 described by Ekstein (1944) since d_{21} and d_{23} have opposite signs.

ACKNOWLEDGMENTS

The author wishes to thank Miss S. Rodwell for assistance in measurements and calculations.

Acknowledgment is made to the Engineer-in-Chief of the General Post Office for permission to make use of the information contained in this paper.

REFERENCES

- BECHMANN, R., 1940, *Hochfrequenztech. u. Elektroakust.*, **56**, 14; 1941, *Z. Phys.*, **117**, 180; 1942, *Ibid.*, **118**, 515, 120, 107; 1949, *Nature, Lond.*, **164**, 190.
 BECHMANN, R., and LYNCH, A. C., 1949, *Nature, Lond.*, **163**, 915.
 CADY, W. G., 1946, *Piezoelectricity* (New York: McGraw-Hill).
 EKSTEIN, H., 1944, *Phys. Rev.*, **66**, 108.
 HIGHT, S. C., and WILLARD, G. W., 1937, *Proc. Inst. Radio Engrs.*, **25**, 549.
 HOLMAN, W. F., 1909, *Ann. Phys., Lpz.*, **29**, 160.
 JAFFE, H., 1948, *Phys. Rev.*, **73**, 1467.
 LYNCH, A. C., 1950, *Proc. Phys. Soc. B*, **63**, 323.
 MÄHLI, H., and TRÖSCH, A., 1947, *Helv. Phys. Acta*, **20**, 253.
 MASON, W. P., 1942, *Electromechanical Transducers* (New York: D. van Nostrand); 1946, *Phys. Rev.*, **70**, 705; 1947, *Proc. Inst. Radio Engrs.*, **35**, 1005.
 MEYER, G., 1937, *Ph.D. Thesis*, Göttingen.
 SPITZER, F., 1938, *Ph.D. Thesis*, Göttingen.
 TAMARU, T., 1905, *Phys. Z.*, **6**, 379.
 1948, *Analytical Chemist*, **20**, 491.

Some Measurements of the Resistivity of Good Insulators

By N. W. RAMSEY

Radiotherapy Department, Charing Cross Hospital

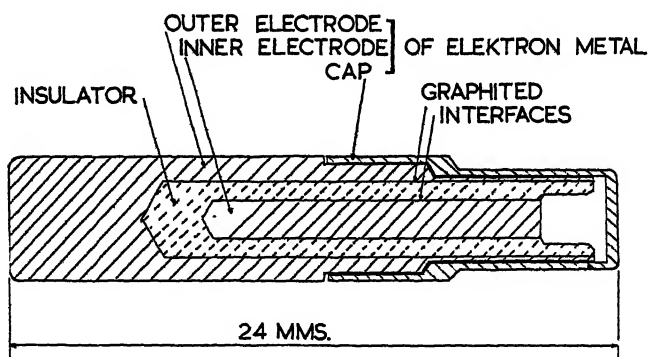
Communicated by J. E. Roberts; MS. received 25th July 1949, and in amended form 3rd January 1950.

ABSTRACT. Measurements of the resistivity of good insulators were carried out by measuring the loss of charge on a condenser. A knowledge of the dimensions of the insulator is not required. It was found that the insulation resistances of amber, alkathene, distrene and Perspex increase with time over a period of weeks, the final values being considerably higher than previously published figures. There is no significant difference between the values for the first three materials. The leakage time necessary for reasonable accuracy is indicated.

§ 1. INTRODUCTION AND EXPERIMENTAL PROCEDURE

FROM experience in Radiological Physics, doubts of previously accepted values had arisen in view of the performance of the type of condenser ionization chamber used in measurements.

Observations were taken of the insulation resistance of the insulators of some condenser ionization chambers of the type used for the measurement of intracavitary x-ray dosage in radiotherapy (Farmer 1945). A cross section of the cylindrical chambers used is seen in the Figure. The insulators under test (see



Condenser ionization chamber.

list below) were coated with graphite inside and outside before assembly, in order to improve contact with the electrodes and to eliminate stray capacities.

Natural amber, two samples.

Alkathene, samples 1 and 2: normal grade 20 alkathene; sample 3: grade 20 made in presence of small percentage of carbon monoxide. All three samples were supplied by Imperial Chemical Industries Ltd.

Distrene, one sample: commercial brand of polystyrene, supplied by BX Plastics Ltd.

Perspex, one sample: unplasticized Perspex acrylic rod (methyl methacrylate monomer), supplied by Imperial Chemical Industries Ltd.

The condensers, after being charged, were kept in a desiccator and the loss in charge over periods varying from a few days to several months was measured with a valve electrometer of the type devised by Farmer for use with small ionization chambers (Farmer 1942).

The potential gradient across the insulators was about 130 volts/mm. when the condensers were freshly charged, except in the case of the Perspex insulators, when the gradient was about 100 volts/mm. During the leakage periods the decrease in potential gradient was up to 50% for Perspex and up to about 20% for the other insulators.

§2. CALCULATION OF RESULTS

Treating the specimen for which the resistivity is to be found as a condenser in parallel with a high resistance, the voltage across the condenser at any time after the condenser was charged is given by

$$V_t = V_0 \exp(-t/CR),$$

where V_0 is the voltage to which the condenser was initially charged, V_t the voltage across the condenser after a leakage time t , C the capacity of the condenser and R the leakage resistance of the condenser.

For a rectangular block of insulation, the resistance between one pair of opposite faces is $R = sd/A$, where s is the resistivity, d the distance of separation between the two faces and A the area of these faces.

The capacity between the same pair of opposite faces is $C = KA/4\pi d$, where K is the dielectric constant.

$$\text{The product } CR = \frac{KA}{4\pi d} \cdot \frac{sd}{A} = \frac{Ks}{4\pi}.$$

Since the last expression is independent of the dimensions of the insulator, it can be used to replace CR for a condenser of any shape. Substituting in the formula for the voltage drop due to leakage, we have

$$V_t = V_0 \exp(-4\pi t/Ks),$$

from which we find the resistivity

$$s = \frac{4\pi t}{K \ln(V_0/V_t)}.$$

Thus all that is needed in order to find the resistivity is a measurement of the voltage across the condenser before and after the leakage period, and a knowledge of t the leakage time and K the dielectric constant. Tedious and possibly inaccurate measurements of insulator dimensions can thus be avoided. It is interesting to note that since the voltage loss of the condenser is independent of its dimensions, no improvement to shorten the long leakage periods can be effected by altering the shape of the insulator.

§3. ACCURACY OF RESULTS

The dielectric constant K can be obtained from published tables with more than sufficient accuracy. The leakage time is also known with ample accuracy, since the leakage periods used were of the order of days or months and could be timed to the nearest hour if necessary. The main cause of error was a variation in the calibration of the valve electrometer between the beginning and end of a

leakage period, sometimes an interval of several months. For this reason the accuracy is often low, particularly where the drop in potential on a condenser is small, but the final results are probably not in error by more than a factor of 3 or 4.

The leakage period necessary to obtain a reasonably large drop in potential on a condenser can readily be estimated if the time constant of the condenser is known. The capacity of a typical condenser is about 3 pF. and the resistance of a specimen having a resistivity of 10^{21} ohm cm. is about 7×10^{19} ohm, so that the time constant in this case is 2.1×10^8 seconds which amounts to 2,430 days or between 6 and 7 years.

§ 4. RESULTS

The results are set out in Table 1. It will be noticed that in general the measured values of resistivity rose during the early stages of the experiments, no further significant increase being detected after a period of approximately 100 days. This rise in resistivity might be attributed either to a long-term dielectric absorption, or to the removal of moisture present initially on the surface of the insulators, as all the condensers were freshly made at the beginning of the experiments.

Table 1

Material	(1)	(2)	(3)	Material	(1)	(2)	(3)
Amber :	5	5	0.61	Alkathene :	1	1	1.9
Sample 1	730	620	14	Sample 3	6	5	2.0
	867	137	14		33	19	5.0
	1023	156	8.1		716	620	14
	1066	43	22		853	137	55
	1248	182	9.6		1009	156	14
					1234	182	20
Amber :	3	3	0.34	Distrene	5	5	1.0
Sample 2	7	4	0.31		28	23	3.6
	14	7	2.62		47	19	2.2
	96	35	0.75		730	620	7.6
	161	65	8.2		867	137	18
Alkathene :	5	5	1.5		1023	156	15
Sample 1	28	23	10		1248	182	13
	730	620	23	Perspex	1	1	0.016
	867	137	46		2	1	0.023
	1023	156	14		7	5	0.086
	1248	182	19		15	8	0.73
Alkathene :	1	1	0.06		34	19	0.34
Sample 2	2	1	0.26		717	620	2.3
	7	5	8.8		854	137	1.4
	15	8	4.7		1010	156	1.6
	34	19	2.3		1053	43	2.0
	717	620	7.1		1235	182	1.9
	854	137	5.3				
	1010	156	8.0				
	1053	43	9.4				
	1235	182	7.4				

(1) Time from start of test (days); (2) leakage period over which loss of charge was measured (days); (3) resistivity in ohm cm. $\times 10^{20}$.

Further tests were made on two batches, each of six condensers, with amber insulators, one batch being kept in air, and the other in a desiccator. No significant difference between the final values of the resistivity for the two groups of insulators was found, beyond that attributable to variations from one specimen to another, and experimental error. All the specimens showed a final resistivity between 10^{19} and 10^{21} ohm cm. Thus it would appear that the rise in resistivity is not due to the drying of the insulator surfaces.

It must be emphasized that the results obtained give low limits for the resistivities, because no correction has been made for the loss of charge due to leakage, if any, over the end surface of the insulators, nor for the ionization produced in the air volume of the associated ionization chambers by stray penetrating radiation. Taking the ionization produced by the latter as 10 ions/cm³/sec., the loss of charge due to this effect is estimated at not more than 1% of the loss due to insulation leakage in any of the condensers.

Some examples of previously published figures are given in Table 2.

Table 2

Material	Volume resistivity	Reference
Amber	5×10^{18} ohm cm. at 22° c.	<i>Handbook of Chemistry and Physics</i> 1944.
Distrene (Polystyrene)	10^{20} ohm cm.	Wearmouth and Cozens 1942.
Alkathene (Polyethylene)	3×10^{17} ohm cm.	I.C.I. 1945.
Perspex (Polymethyl- methacrylate)	$>10^{18}$ ohm cm. at 20° c.	I.C.I. 1944.

§ 5. DISCUSSION OF RESULTS

It will be seen that the magnitudes of the resistivities found are, in most cases, higher than figures previously published by a factor of the order of 10^4 . No attempt has been made to carry out the measurement by a direct current method for the purpose of comparison. In § 3 it is shown that a condenser having a dielectric with a resistivity of 10^{21} ohm cm. has a time constant of 2,430 days or 6 to 7 years. It is common experience that condenser ionization chambers in normal use have a time constant considerably greater than 0.24 day, which is the value corresponding to a dielectric having a resistivity of 10^{17} ohm cm.

The results quoted in Table 2 for alkathene and Perspex were obtained by measurement of the direct current through a specimen, after an electrification period of 1 minute. The fact that these figures are so much smaller than those in Table 1 might well be taken as additional evidence that the current through an insulator decreases with time after the application of voltage.

It appears that, for electrometer work, amber is no better than the newer insulating materials; from a practical point of view the latter are more readily obtainable, and easier to fabricate. However, for applications in which the insulation is subjected to irradiation by x-rays, amber is far superior to the synthetic materials in that it will withstand a greater amount of x-radiation before the resistivity is appreciably lowered.*

* It is hoped to publish some information on this later.

The reason for the existence of any current flow in good insulators has for a long time been the subject of speculation. It is very often attributed to the presence of impurities, in particular, small amounts of moisture. The flow of current through the specimens tested was of the order of $110 \text{ ions/cm}^3/\text{sec.}$, when the resistivity had reached its final level, and it is conceivable that the initial rise in resistivity might be caused by the gradual dissipation by electrolysis of minute traces of moisture in the specimen. It is not obvious on this assumption, however, why the conduction current should attain an approximately steady value after the initial period of increase. This might be expected in those specimens maintained in normal moist air, but not in those which were kept dry.

A possible explanation of the existence of this final steady current is the production of ionization in the insulators by penetrating radiation. Some support for this theory can be found in the fact that the final values of conduction current are of the same order in three of the materials tested, amber, alkathene and distrene, and only a factor of 10 higher than this in Perspex. Taking rough values for the densities of air and the insulators as 0.001 and 1 gm/cm^3 respectively, and neglecting differences between their atomic numbers, it would be expected that something like $10,000 \text{ ions/cm}^3/\text{sec.}$ would be produced in the insulators, by penetrating radiation, if the corresponding figure for air is $10 \text{ ions/cm}^3/\text{sec.}$ as assumed above. With the potential gradients used in the tests of 130 volts/mm. or so, a large measure of ionic recombination would be expected, but the figure for the flow of current of $110 \text{ ions/cm}^3/\text{sec.}$ quoted above could reasonably be attributed to this cause.

ACKNOWLEDGMENT

I am indebted to Professor J. E. Roberts of the Barnato-Joel Laboratories, Middlesex Hospital, where these measurements were mostly carried out, for his continuing interest and encouragement during long periods of waiting.

REFERENCES

- FARMER, F. T., 1942, *Proc. Phys. Soc.*, **54**, 435; 1945, *Brit. J. Radiology*, **18**, 148.
Handbook of Chemistry and Physics, 1944, 28th Edition (Cleveland, Ohio: Chemical Rubber Publishing Co.).
IMPERIAL CHEMICAL INDUSTRIES, 1944, *Plastics Technical Bulletin*, No. 10; 1945, *Ibid.*, No. 13.
WEARMOUTH, W. G., and COZENS, E. G., 1942, *Electronic Engng.*, **14**, 667.

A Radio Meteorological Investigation in the South Island of New Zealand

BY B. MILNES* AND R. S. UNWIN

Departments of Scientific and Industrial Research, United Kingdom and New Zealand,
respectively.

*Communicated by R. L. Smith-Rose; MS. received 15th June 1949, and in amended form
2nd February 1950*

ABSTRACT. A Föhn wind blowing off-shore produces super-refraction at radio frequencies. This results in the guiding of very high frequency radio waves beyond the horizon in a radio duct close to the sea surface. The problems involved and the experimental techniques adopted in a long term investigation of this phenomenon are described, and selected data are discussed. Field observations are presented in the form of isopleths of the relevant meteorological quantities and of modified refractive index over the area under investigation, together with contours of radio field strength for wavelengths of 300 cm., 60 cm., 10 cm. and 3 cm., corresponding to the particular meteorological situations described.

It is found that with off-shore winds of strength up to force 6 radio ducts of about 100 m. height are formed and the field strength everywhere beyond the horizon is much increased. In the duct itself, at ranges up to 200 km. from the coast, field strengths approximately equal to the free space level are obtained on wavelengths of 10 cm. and 3 cm., and up to 10 or 20 db. below this level on a wavelength of 60 cm. The duct is not sufficiently high to trap radiation at a wavelength of 300 cm. With stronger winds turbulence limits the duct height to 20 m. and radiation of wavelength above 10 cm. is no longer trapped. The presence of a local sea breeze underneath the off-shore drift has little effect on propagation normal to the coastline.

§ 1. INTRODUCTION

(i) *The Phenomenon of Super-refraction*

WHEN a warm dry wind blows over a relatively cool sea, steep gradients of temperature and humidity occur in the lower atmosphere. Such conditions produce an abnormally great lapse rate of refractive index with consequent super-refraction of radio waves. Super-refraction results in a concentration of energy at low levels in a so-called radio duct and this causes the field strength beyond the horizon to be very much increased. The structure and properties of radio ducts were intensively studied during the war and the state of knowledge of the whole subject up to 1946 has been summarized in the Report of the Physical Society and Royal Meteorological Society Conference (1946) and the Report of the Committee on Propagation, N.D.R.C. (1946).

The extent and intensity of an advection duct depend upon the initial temperature difference between the air and the sea, the moisture content of the air before it reaches the sea, and the wind speed and direction at various heights. Further, for a given set of conditions the amount of energy trapped in a radio duct falls off rapidly with decrease in frequency below a critical frequency which is the cut-off of the lowest frequency mode of the duct. Moreover the duct is defined in terms of the distribution of refractive index of the air which can be calculated from a knowledge of the temperature, humidity and pressure of the air (Englund, Crawford and Mumford 1935). Theories have been developed so that, knowing

* Now at Telecommunications Research Establishment, Ministry of Supply, Malvern.

the distribution of refractive index in the lower troposphere, the amount of energy trapped within the duct may be calculated. Booker and Walkinshaw (1946) have developed a waveguide theory to provide approximate solutions to the problem of calculating the field distribution inside and above the duct, and Price (1948) has described methods based on ray tracing to achieve the same ends. J. E. Freehafer working at the Massachusetts Institute of Technology (U.S.A.) has also developed the ray-tracing technique. Experiments made in many parts of the world (Katzin, Bauchman and Binnian 1947, Megaw 1946, Smith-Rose and Stickland 1946) have provided data with which to check these theories, but, especially in the case of coastal super-refraction, the meteorological data have usually been too meagre to permit of reliable correlation with the radio observations, or to provide satisfactory checks on the current theories of diffusion as applied to an air mass passing off-shore.

The 'Canterbury Project' was initiated jointly by the Telecommunications Research Establishment, Ministry of Supply, and the Departments of Scientific and Industrial Research, United Kingdom and New Zealand, with some material assistance from the United States of America, to study the whole problem and develop if possible a suitable forecasting technique for the onset, intensity and duration of super-refractive conditions.

This paper is confined to a description of the experimental techniques adopted to obtain the necessary meteorological and radio data, and a discussion of selected results which are typical of the super-refractive conditions experienced.

(ii) *Selection of Site for an Investigation*

In planning an experimental investigation of the problem a number of aspects must be considered. It is necessary to make detailed radio field strength and meteorological observations over a wide area both within and beyond the horizon and within and above the radio duct. To correlate radio and meteorological data ideally an instantaneous picture is required over the whole area, but to trace the modification of an air mass passing off-shore a number of measurements following each other in time and space are desirable. These conflicting aspects can be resolved if a steady meteorological situation can be found, where measurements over a comparatively short period of time (say two hours), will give a reliable picture of the radio propagation conditions and the modification of the air mass. The problem can be further reduced to two dimensions where an air mass moves off-shore perpendicular to a long straight coastline. In choosing a site irregularities of the land surface should be as few as possible, there should be no obstructions off the coast, and, of course, suitable weather conditions (i.e. the off-shore movement of homogeneous air) should be frequent.

This wide set of conditions is almost ideally satisfied in mid-Canterbury on the east coast of the South Island of New Zealand (Figure 1). In this region the prevailing weather is westerly, and the well-known north-west wind of Canterbury blows frequently all the year round, lasting several days at a time. It is especially warm and dry during most of the year, due to the Föhn effect produced by the Southern Alps, which extend up to an average height of about 2,000 metres approximately at right angles to the prevailing wind direction. The Canterbury Plains are about 60 kilometres wide at Ashburton, which was chosen as the Headquarters area for the investigation, with a gentle almost uniform slope towards the sea, free from hills and other major obstructions. To seaward the area is

clear of islands. With the wind development at right angles to the coastline this region forms an almost ideal natural laboratory for investigating the formation of radio ducts off-shore with accompanying super-refraction at very high frequencies. The map (Figure 1) shows the line along which observations were made.

(iii) Measurements Required

From the foregoing it will be clear that detailed measurements of temperature, humidity, wind speed and direction of an air mass passing off-shore are required over land and sea, out to a considerable distance from the coast. A series of vertical soundings up to 600 m. or so should be made over a limited time interval simultaneously at various points along the line of operations: Under steady conditions the data obtained will yield information showing the modification of an air mass passing off-shore, and give the distribution of modified refractive index* over the sea showing the radio duct in full. For the development of a forecasting technique for super-refraction, continuous records of pressure, temperature, humidity and wind speed and direction over land are required, as well as synoptic charts of the whole area round New Zealand.

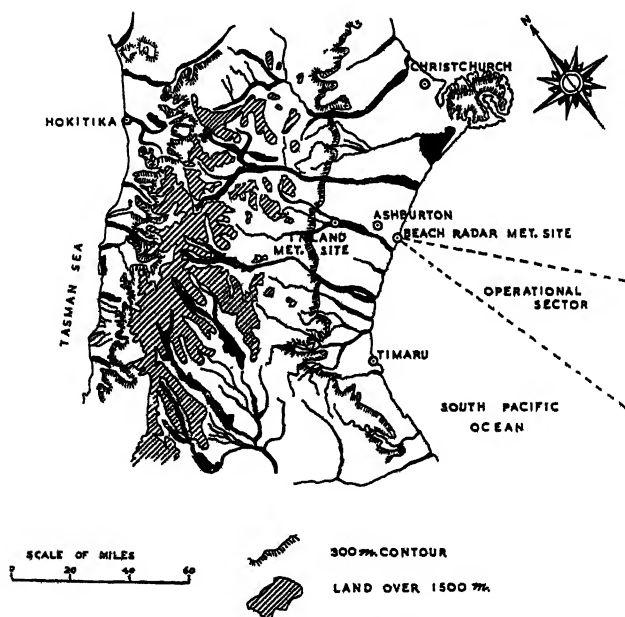


Figure 1. Sketch map of Canterbury and Westland, showing area of operations.

Super-refraction is most marked at microwave frequencies, but for a thorough investigation simultaneous measurements distributed over a wide frequency band from say 100 Mc/s. up to 10,000 Mc/s. should be made. Receivers at the selected frequencies must be placed at or near the coastline, preferably at different heights above sea level, and at the same time as the meteorological measurements are

* The curvature of a ray relative to the surface of the earth, at a height h , is given by $(d\mu/dh) + (1/a)$, where a is the radius of the earth, therefore by adding the term h/a to the ordinary refractive index μ we obtain a fictitious refractive index which automatically takes account of the earth's curvature in any calculations of ray paths. Since the refractive index of air normally exceeds unity by only a few parts in 10^6 a more convenient unit (the Munit) is obtained by multiplying this excess by 10^6 . Thus we have modified refractive index $M = 10^6(\mu - 1 + h/a)$.

being made, observations of the signal strength from transmitters situated at various heights within and above the duct at selected ranges from the coast are required. These observations should be sufficiently closely spaced to provide a complete picture of the distribution of field strength within and above the duct out to the maximum range covered.

§ 2. EXPERIMENTAL PROCEDURE

(i) General

In order to meet the requirements stated in § 1 (iii), the only practical possibility was to use aircraft to obtain the necessary rapid coverage. Over the sea the aircraft observations were supplemented by observations from a trawler, and over land by observations from three mobile sounding stations. The sounding trucks and the trawler were fitted with wired sonde equipment whereby rapid and frequent soundings could be made up to heights of between 150 and 600 m. depending on the wind conditions.

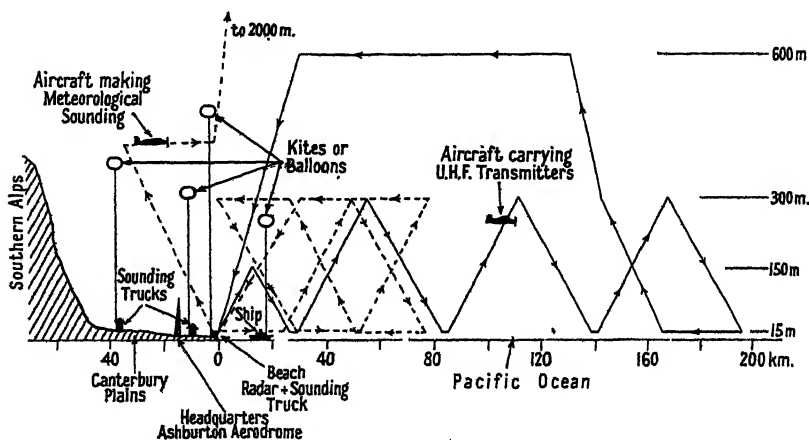


Figure 2. Cross section of area of operations, showing participating units.

The aircraft carried, in addition to meteorological instruments, transmitters operating on wavelengths of 300 cm., 60 cm., 10 cm. and 3 cm., and the signals from these transmitters were measured by the appropriate receivers situated at the coast. 10 cm. and 3 cm. receivers were located at heights of 8 m. and 26 m. above mean sea level, this duplication of receivers enabling any variation of trapping with height of source to be studied. The 300 cm. and 60 cm. receivers were at heights of 11 m. and 27 m. respectively. A higher receiving site would have been desirable near the top of, or just above, the average duct height experienced, but this, unfortunately, was not available. Horizontal polarization was used on all wavelengths.

Standard meteorological instruments provided a continuous record of surface temperature, pressure, humidity and wind speed and direction at stations at the coast, 15 km. inland and 38 km. inland along the line of observations.

Figure 2 depicts a cross section of the area looking from the south, and shows the general arrangement of the experiment. The Headquarters at Ashburton Aerodrome was in radio-telephone or land-line communication with all observing posts, and one or two persons controlled all observations from there. Land

soundings were usually started one to three hours earlier than an aircraft flight to check the homogeneity of the air mass. If two aircraft were used simultaneously, only one provided signals for radio measurements, and the other, making meteorological soundings only, concentrated on the region up to 60 or 70 km. from the coast where the modification of the air mass is most rapid. For the same reason the ship usually remained within 30 km. from the coast, so that the very sharp gradients occurring in this region could be examined in detail, but it made periodical trips out to 150 km. from the coastline for measurement of sea temperature.

It will be seen from Figure 2 that, with the three sounding posts on land, the trawler off the coast, and the numerous ascents and descents of both aircraft up to 300 m. or higher, a wealth of meteorological and radio data could be obtained in the space of about two hours. On most occasions isopleths of temperature, humidity, modified refractive index and field strength on all four wavelengths could be drawn over the whole area.

Through the courtesy of the New Zealand Meteorological Office radiosonde ascents were made when desired from Hokitika on the West Coast. These provided data on the properties of the air mass before it crossed the mountain chain of the Southern Alps and, when compared with its properties on the eastern side, could yield valuable information on the Föhn effect, the cause of the warmth and dryness of the north-west wind in Canterbury.

(ii) *Meteorological Equipment*

The main requirement of the meteorological equipment was that it should have a rapid response to small changes in temperature and humidity since it was the gradients of these quantities with height that were the primary consideration; in addition the equipment had to be rugged and simple in operation for use in the field and on board ship or aircraft.

(a) *Aircraft installation.*

The aircraft installation consisted of a 'spirit in glass' wet and dry bulb psychrometer. Temperatures could be read to $\pm 0.2^\circ\text{F.}$, giving a maximum possible error in modified refractive index of ± 1.3 Munits at 50°F. and ± 2 Munits at 80°F. The lag of the instrument was 8 seconds at an airspeed of 185 km/hr. since the rate of ascent and descent of the aircraft did not exceed 40 metres per minute, this lag introduced only a second order error except in the sharpest gradients. These sharp gradients were usually found close to the sea surface, so that when the finish of a descent and the start of a subsequent ascent were close together in range, errors due to lag could be smoothed out when necessary. Height was measured by an aneroid or radio altimeter, with a probable error between +6 and -3 m. in the first hundred metres. Height differences could be relied on to ± 5 m. or better.

The psychrometer was read every 15 to 30 seconds, corresponding to a height interval of 10 to 20 metres depending on the sharpness of the gradients encountered; airspeeds were also noted. Altitude was logged by the navigator every 30 seconds.

(b) *Wired sonde equipment.*

The wired sonde equipment used in the sounding trucks and trawler followed the technique developed by P. A. Anderson, G. L. Barker, K. E. Fitzsimmons and S. I. Stephenson at Washington State College (U.S.A.). Temperature measurement could be relied on to $\pm 1^\circ\text{F.}$ or better, and relative humidity to

$\pm 3\%$ in widely varying conditions such as occurred in a sounding through a sea breeze into a dry north-west wind. This gave a possible error in modified refractive index of ± 3 Munits at 50°F. and ± 8 Munits at 80°F. , but in a more normal situation, with smaller gradients of humidity, the error would be less than half this amount.

Detailed measurements of wind gradient over the land were made with a small cup-type anemometer hoisted aloft by kite or balloon. This wind information was supplemented by frequent wind-finding runs made at the Headquarters site using standard pilot balloon and theodolite technique.

In winds of strength up to force 3 on the Beaufort scale the equipment (sonde elements or anemometer) was hoisted by balloon, in winds of force 3 to 5 by Seyppang kite and balloon combination, and in stronger winds by the kite alone.

(iii) *Radio Equipment*

Since the signals to be measured were from a moving source, an airborne transmitter, it was essential to use pulse transmission in order to provide range information.

The four transmitters in the aircraft were triggered by signals received from one of the 10 cm. ground stations, and the signal strength of the returned pulses from these transmitters was measured at the ground stations by visual observation on an 'A' scope display, i.e. amplitude-range display. On the 300 cm. and 60 cm. receivers a comparator pulse was injected from a signal generator in parallel with the incoming radio signal. This comparator pulse was maintained at the same amplitude as the signal received from the aircraft by manual control of the calibrated signal generator attenuator, from which the signal strength could be read directly in microvolts or in decibels above some arbitrary value. On the microwave receivers, where the signals varied more rapidly, the amplitude was estimated directly against a series of horizontal lines engraved on the face of the display tube.

The signal strength was observed synchronously on all six receivers every ten seconds throughout an aircraft flight, i.e. a reading was obtained at average height intervals of 6 m. on each inclined run. Usually the aircraft made about eight ascents and descents between 300 m. and 15 m. out to a range of 180 km., as well as a level run for 100 km. at a height of 600 m. during one flight, hence a comprehensive picture of the radio field strength pattern was obtained over the whole area.

The level run towards or away from the ground stations at a height of 600 m. was made on almost every occasion primarily to enable a value of the 'free-space' signal strength for the various R.F. channels to be established for each operation. On this run the aircraft passed through a number of the normal interference lobes produced by reflection from the sea surface, and a $1/r^2$ attenuation curve could usually be fitted to the peaks of these lobes, when plotted, correct to one or two decibels. From this curve, by assuming a value of -1 for the reflection coefficient of the sea, the free-space signal at any range could easily be calculated. By obtaining a free-space signal level in this way it was not necessary to maintain an accurate check on the absolute sensitivity of the ground station receivers or the power output of the airborne transmitters. It is estimated that the signal strength observations on 300 cm. and 60 cm. are accurate to about 1 db. and on the 10 cm. and 3 cm. wavelengths to 2 or 3 db.

The transmitters used in the aircraft were modified airborne radar transmitters and all four antennae had wide coverage in the horizontal plane to reduce aspect

effects. The 10 cm. receiver in the aircraft supplied triggering pulses to all four transmitters, locked in phase to the received pulses from the 10 cm. ground transmitter, and this link was such that it remained operative down to about 35 db. below the free-space signal at a range of 150 km.

The ground station receivers were of standard radar type, all receiver displays being locked in phase with the 10 cm. transmitter providing the triggering pulse. Manual control of the antennae and tuning was used on all the ground stations.

§ 3. DISCUSSION OF SELECTED DATA

Observations were begun in September 1946 and continued until the end of 1947, by which time 84 sets of observations had been obtained on 70 days. In addition, 13 sets of observations confined to the sea breeze structure and without radio data were obtained. A sufficient number of 'nor'westers' occurred throughout the whole period for the seasonal variations to be established, and the conditions encountered varied between wide limits. Owing to technical difficulties with the aircraft no observations were made at night.

It is not possible in the space of this paper to present more than a very small fraction of the total data collected, and for further information reference should be made to the complete record of observational data which is at present in course of publication in New Zealand.

In selecting the data for discussion a situation has been chosen in which the refractive index gradient is approximately standard, and two examples of north-westerly conditions have been included, one occurring in early spring and the other in early summer. With a light north-west wind a sea breeze often penetrates underneath the off-shore drift, and an example of this complication has also been included. The remaining situations have been selected because of some peculiarity markedly affecting radio propagation.

In spite of the many advantages of the mid-Canterbury Plains as a site for an investigation, on most occasions there was a sufficient change in the situation over the period covered by an aircraft flight to necessitate its being taken into account when interpreting the data. The effect was usually small, though quite distinct; but in complex situations such as occurred with a sea breeze or north-easterly drift under a 'nor'wester', some smoothing of the observations was necessary, or separate isopleths were required to show the situation at a later time.

The field strength contours for wavelengths of 60 cm. and 300 cm. can usually be drawn with very little difficulty, and present a reliable picture of the situation. The contours of field strength for 10 cm. and 3 cm., however, usually present some difficulty because of greater fluctuations in the observed values, and it is considered that, in many instances, the exact situation shown never existed as an instantaneous picture, but the contours give an indication of the amount of energy leaking from the duct. These points must be borne in mind when examining the data. The contours are in decibels on a scale such that the free-space radiation field at one metre from the transmitter is 150 db. (50 db. at 100 km.).

(i) *Nearly Standard Propagation*

The observations of 17th October 1947 reveal meteorological conditions giving rise to almost standard refractive index gradients (Figure 3), with no trapping of radio energy, and are included in this paper for comparison with the examples of anomalous conditions which cause super-refraction.

Conditions on this day resulted in the formation of an evaporation duct, about 5 m. in height and having 5 Munits inversion, out to a range of about 100 km. Above 5 m. the modified refractive index increased almost linearly with height to give a normal gradient (Figure 3). There was no trapping of energy on wavelengths of 300 cm., 60 cm. and 10 cm., or on 3 cm. where the source was above the duct. A small amount of trapping was noticed on the low-sited 3 cm. receiver on one run only, and this is shown by the dotted line in the figure.

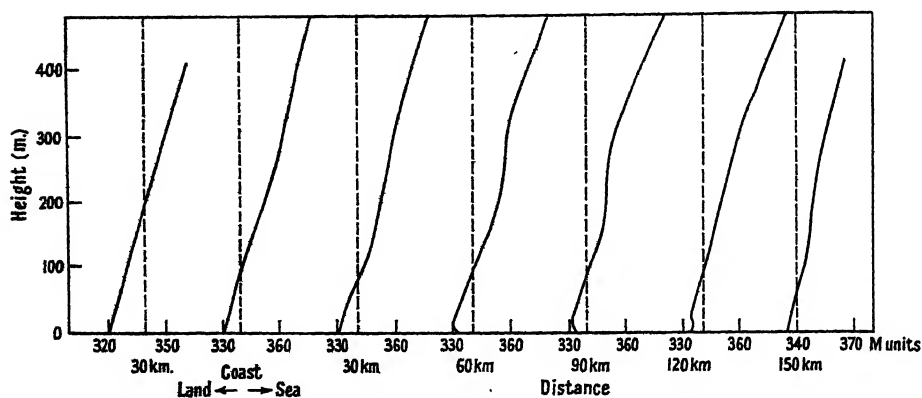


Figure 3. (M, h) curves at various ranges, 17th October 1947, 1130 hours, --- $M=340$. [M at sea surface = 336 approx.]

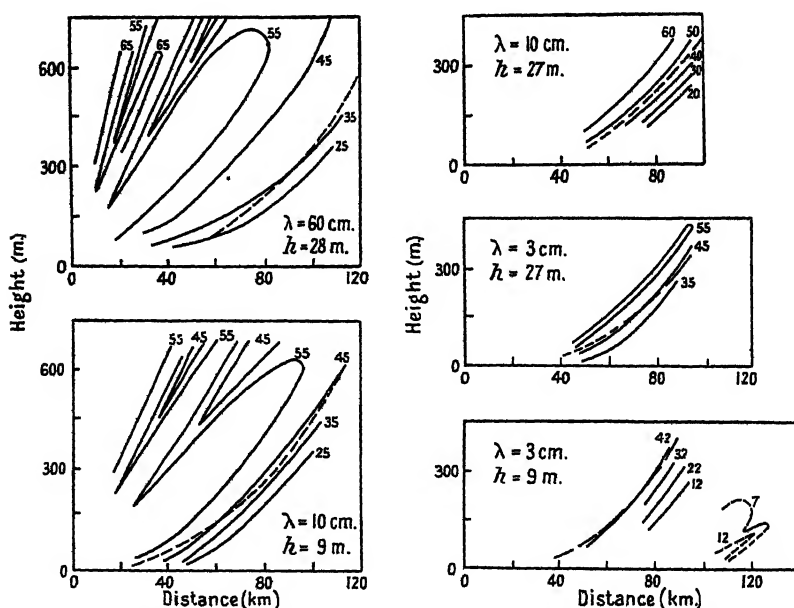


Figure 4. Field strength contours, 17th October 1947, --- horizon line for a standard atmosphere.

The general appearance of the field strength contours (Figure 4) is seen to be the same on all wavelengths. The signal falls off very rapidly below the first interference lobe, i.e. below the horizon, and does not increase again as the sea surface is approached, indicating that the amount of super-refraction, if any, is negligible.

(ii) *An Early Spring 'Nor'wester'*

An example of a typical 'nor'wester' of early spring was that which occurred on 5th August 1947. The synoptic situation (Figure 5) shows an intense ridge of high pressure north of New Zealand extending across the North Tasman Sea and eastwards into the Pacific. The resulting gradient across New Zealand gave a north-west to north-north-west wind of force 3 to 6 blowing off-shore in mid-Canterbury throughout the day.

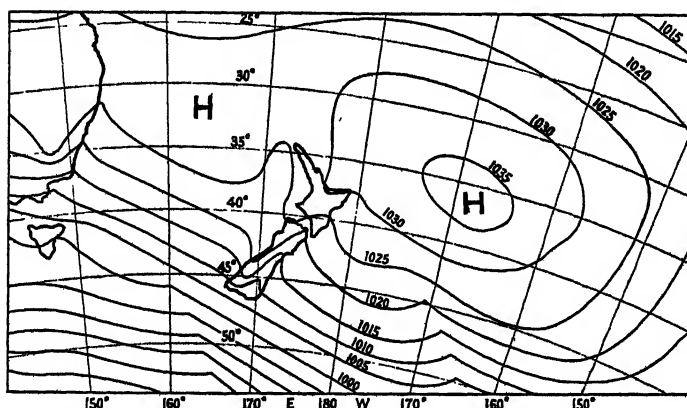


Figure 5. Isobars at 06 hours local time, 5th August 1947.

The potential temperature of the air mass in the early afternoon when the observations were made was about 70°F. , and the specific humidity 4 to 4.5 gm/kg. The sea surface temperature varied between 46°F. close in-shore and 48°F. at about 70 km. from the coast, decreasing again beyond 120 km. to 46.5°F. The corresponding saturation values of specific humidity are 6.6 and 7.0 gm/kg. respectively. Thus the resulting temperature excess and humidity deficit between unmodified air and sea is 22 to 24°F. and between 2 and 3 gm/kg., sufficient to cause a marked advection duct off-shore.

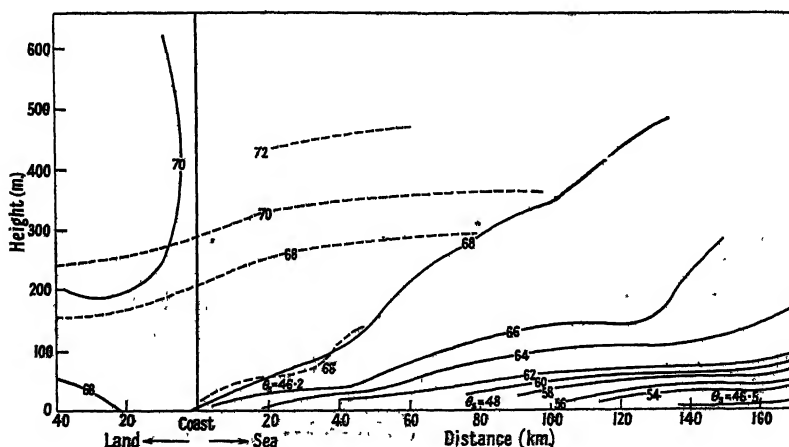


Figure 6. Isopleths of potential temperature (in degrees F.), c. 14 hours 5th August 1947, --- situation at 15 hours.

The modification of the air mass off-shore is typical of a 'nor'wester' at this time of the year. The isopleths clearly depict the cooling and moistening of the air over the sea surface (Figures 6 and 7). However, the dotted lines showing the situation about an hour later than that represented by the full lines show that there is some inhomogeneity in the air mass crossing the coast during the period.

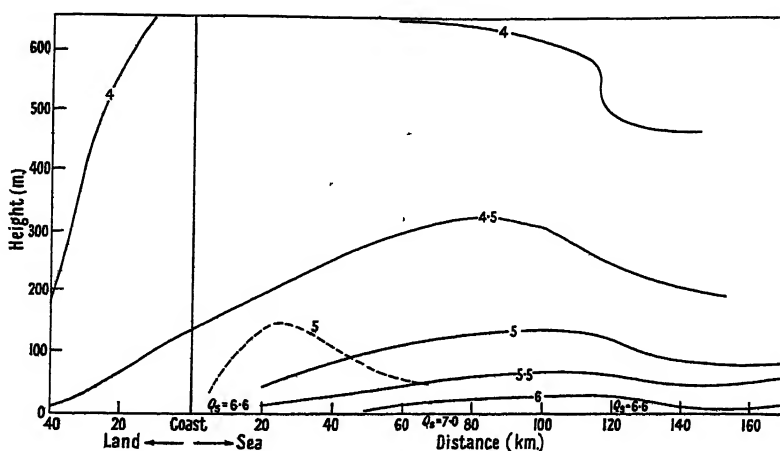


Figure 7. Isopleths of specific humidity (in gm/kg.), c. 14 hours, 5th August 1947, ----- situation at 15 hours.

The distribution of modified refractive index over the area shows a marked duct extending out to the maximum range covered, 160 km. from the coast (Figure 8). The small duct at the beach is due to the modification of the air by the land surface, largely caused by evaporation from the ground, a common feature of the Canterbury 'nor'wester' at this time of the year. The duct close in-shore

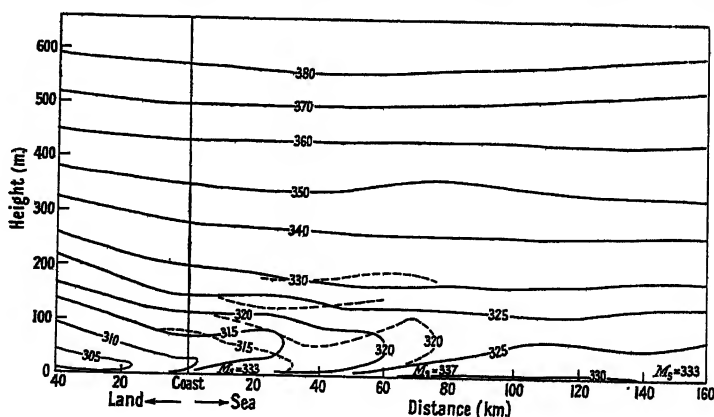


Figure 8. Isopleths of modified refractive index, M , c. 14 hours, 5th August 1947, ----- situation at 15 hours.

is low and intense, gradually decreasing in intensity and increasing in height with increasing distance off-shore. The fall in sea temperature beyond 120 km. is accompanied by a decrease in the surface value of M by 4 units, hence the decrease in intensity of the duct is greater than it would be with constant sea surface temperature. The slight intensification of the duct at 160 km. is due either to the

subsidence of warm upper air, or to inhomogeneity of the air mass. However, the height and intensity of the duct at all ranges, with a maximum height of approximately 100 m. beyond about 120 km. off-shore, is typical of a spring or summer 'nor'wester'. As would be expected, the change is most rapid close to the shore, the variation beyond 80 km. taking place very slowly. At 80 km. the height and intensity are 70 m. and 16 Munits, at 120 km., 90 m. and 11 Munits, at 160 km., 90 m. and 10 Munits (Figure 9).

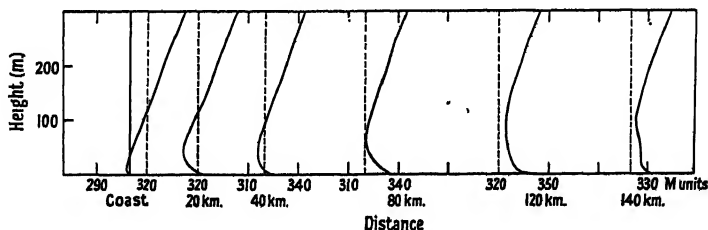


Figure 9. (M, h) curves at various ranges off-shore, c. 14 hours, 5th August 1947, ---- $M=320$.

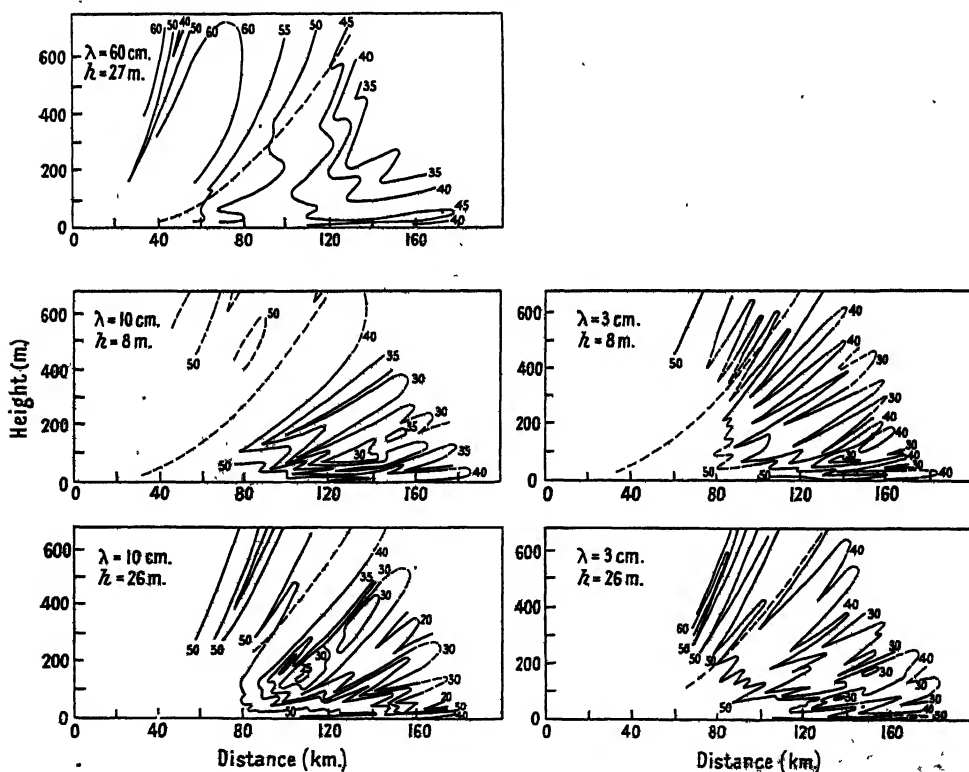


Figure 10. Field strength contours, c. 14 hours, 5th August 1947, ---- horizon line for a standard atmosphere.

Figure 10 shows the field strength contours obtained on this day on wavelengths of 60 cm., 10 cm. and 3 cm. The contours show that there is considerable trapping of energy in a narrow channel close to the sea surface on all three wavelengths. On 10 cm. and 3 cm. the strongest signal in the duct occurs between 15 and 25 m. from the surface and on 60 cm. between 25 and 45 m.

The amount of trapping increases with decrease in wavelength; at a range of 160 km. the peak signal in the duct is 4 db. below the free-space signal on 60 cm., 2 to 3 db. above the free-space value on 10 cm. and 6 or 7 db. above the free-space value on 3 cm.

The contours obtained from the high and low sited receivers on the same wavelength show a somewhat similar structure below the horizon line and give equal peak signals in the duct relative to the free-space value, within the limits of experimental error.

In the diffraction region, between the duct and the interference lobes, leakage from the duct has resulted in a complicated field pattern, especially for 10 cm. and 3 cm., with signal strengths which, although above normal, are nevertheless about 10 db. below the level of the duct signal. It should be noted that the contours are for one-way transmission only, therefore in considering radar coverage the difference in decibel value between one contour and the next would be double that shown. The energy leaking from the duct would then assume far less importance and could generally be neglected for practical purposes. The energy found within the duct constitutes the greater portion of the energy refracted out of the normal lobe pattern.

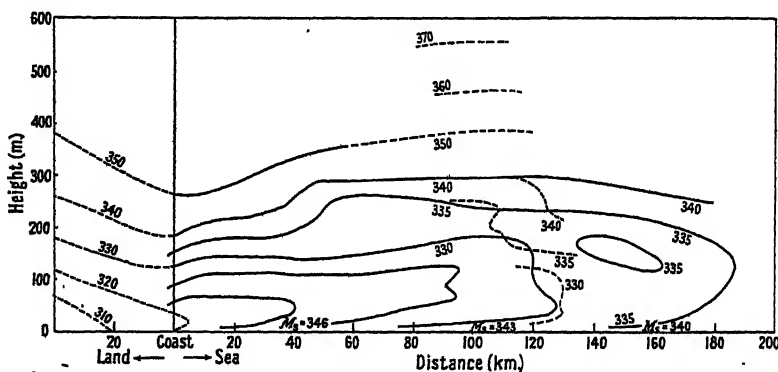


Figure 11. Isopleths of modified refractive index, M , c. 1130 local time, 3rd November 1947, ---- situation c. 1215.

(iii) *An Early Summer 'Nor'wester'*

On 3rd November 1947 a 'nor'wester' of force 3-4 blew off-shore in mid-Canterbury throughout the day. The temperature excess, 16-22°F., and humidity deficit, average 3.5 gm/kg., were not very different from the values for 5th August, and we should therefore expect the resulting situation to be similar.

The isopleths of modified refractive index and the derived (M, h) curves (Figures 11 and 12) are essentially very similar to those of 5th August, the main differences being the much smaller modification over land at this time of the year, and the greater height of the duct due to the slightly greater humidity deficit and the more rapid modification of the air with the smaller wind velocity. At 80 km. the duct is about 90 m. in height with an intensity of 21 Munits; beyond this range it becomes higher but weaker due mainly to a decrease in sea temperature. At 150 km. it is 11 Munits to 150 m. and at 180 km. 7 Munits to 120 m.

The field-strength contours show that the amount of trapping on all wavelengths is much the same as on 5th August, except that the low level peak for 3 cm. is relatively weaker than that for 10 cm. This latter effect is possibly due to

mutual interference between the greater number of propagation modes trapped by this higher duct for the 3 cm. wavelength. Again, owing probably to the greater height of the duct compared with the previous example, the low level peak at 160 km. off-shore for 10 cm. and 3 cm. is between 30 and 60 metres from the surface; for 60 cm. it lies between 45 and 75 m.—which is considerably higher in both cases than on 5th August.

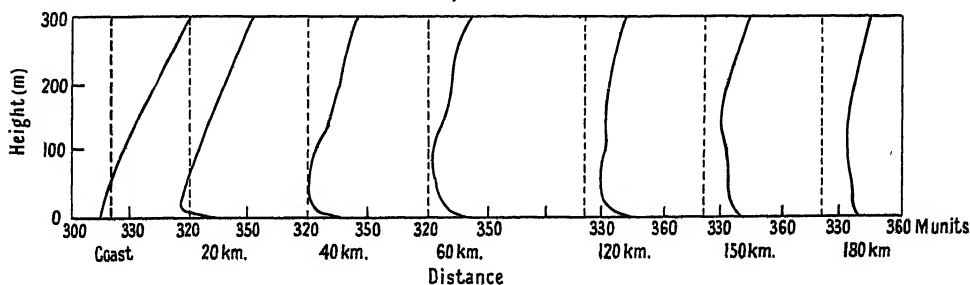


Figure 12. (M, h) curves at various ranges off-shore, c. 1130 local time, 3rd November 1947, ---- $M=320$.

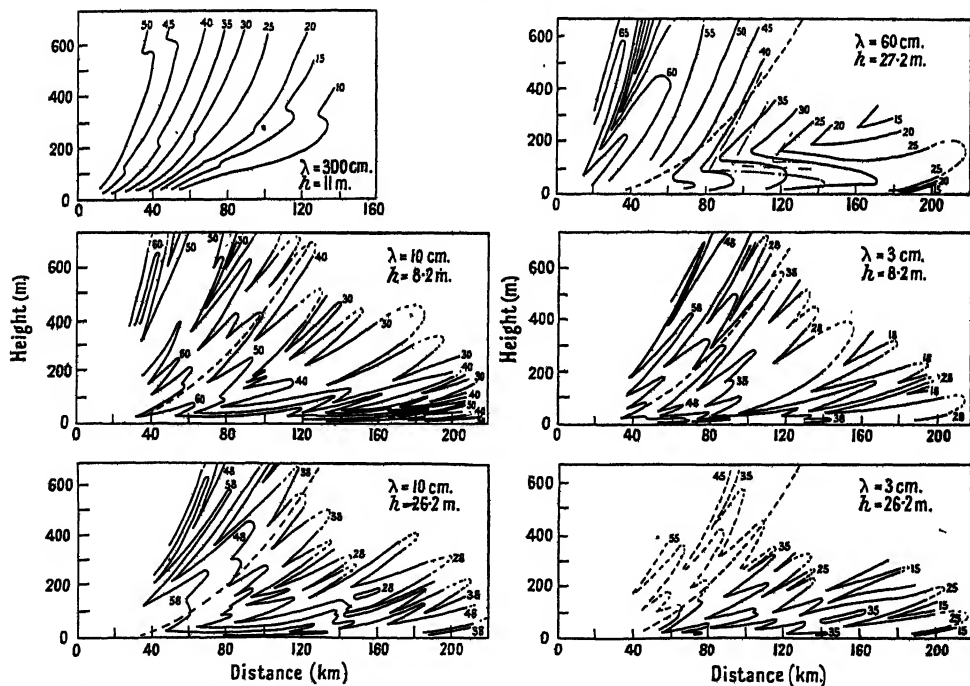


Figure 13. Field strength contours (in db.), c. 1130 local time, 3rd November 1947, ---- horizon line for a standard atmosphere.

The 60 cm. wavelength contours show a particularly simple structure and give a very clear picture of the guiding of energy by the duct round the surface of the earth when plotted on a curved earth diagram (Figure 14).

(iv) Attenuation in the Duct

Aircraft flights were made on several occasions at a constant height in the duct, to study the variation of signal strength in the duct with range. The results of such a flight on 21st October 1947, at a height of 45 m., are illustrated in Figures

15 and 16. There was a strong duct of 20 to 25 M units up to 60 m. in height out to the maximum range of observation, 160 km. This duct height, although insufficient to produce complete trapping at 60 cm., was great enough to cause

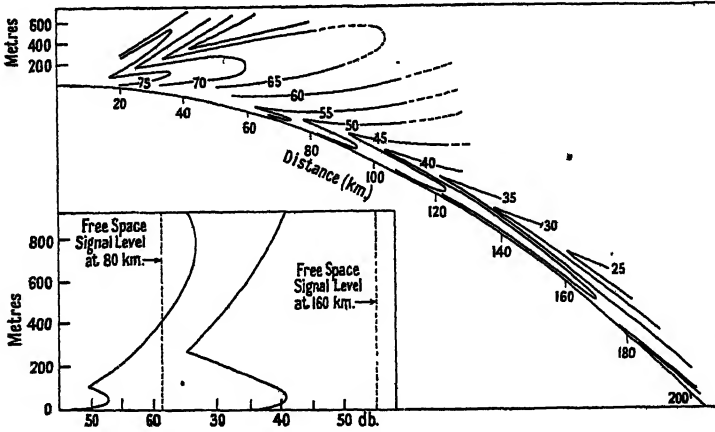


Figure 14. Field strength contours* for wavelength of 60 cm., 3rd November 1947.
Inset: Height-gain curves derived from contours.

* The figures on the contours and in the abscissae of the inset are 10 db. high.

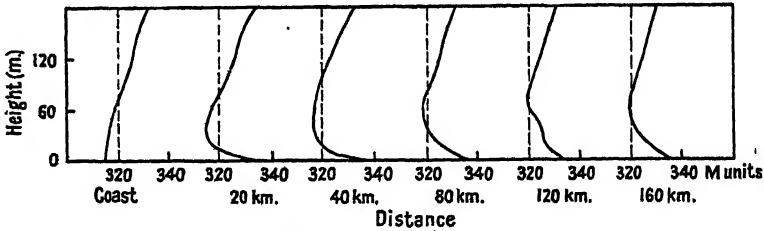


Figure 15. (M, h) curves at various ranges off-shore, 21st October 1947, 11 hours. ---- $M=320$.

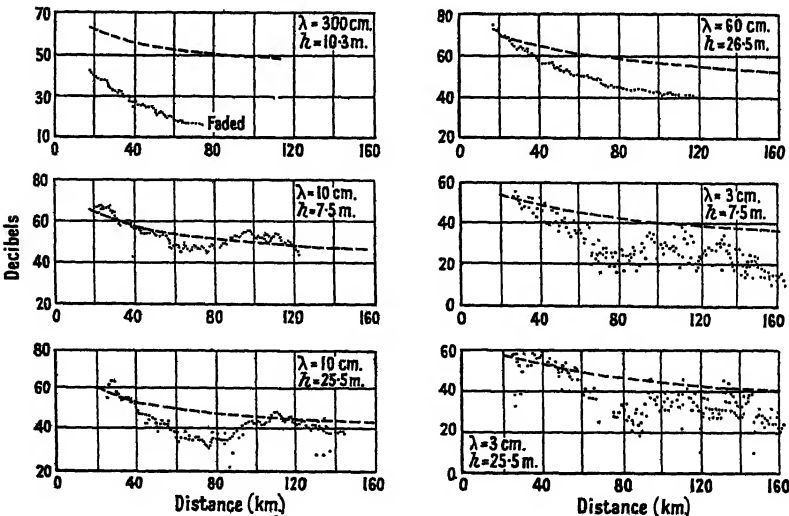


Figure 16. Attenuation in the duct, 21st October 1947. observed values, ---- free-space signal level.

considerable super-refraction at this wavelength and was sufficient for trapping to take place at wavelengths of 10 cm. and 3 cm. The 300 cm. and 60 cm. signals show a steady decrease with increasing range and are always below the free space signal value. Over the range of observations on the 60 cm. channel, 20–120 km., a $1/r^4$ law fits the observed values almost exactly.

For the wavelengths of 10 cm. and 3 cm. the signal oscillates slowly about the free-space value, all four channels showing somewhat similar fluctuations, with a region of weaker signal between about 60 km. and 90 km. The increase beyond this range may have been due to an increase in duct height causing the maximum signal to approach the height at which the aircraft was flying (45 m.). Alternatively, it may have been caused either by slow temporary fading, due to changed meteorological conditions, or by the presence of islands of strong and weak signal due to

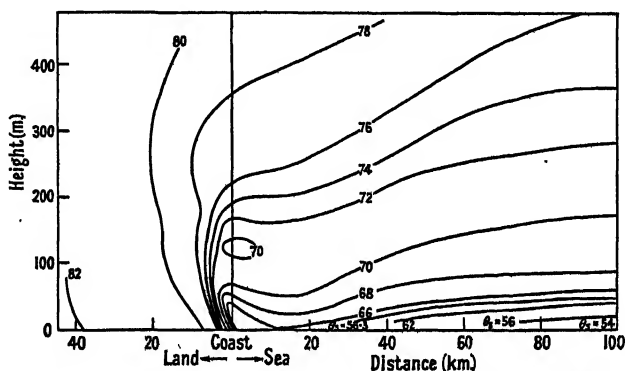


Figure 17. Isopleths of potential temperature (in degrees F.), c. 1430 local time, 8th December 1946.

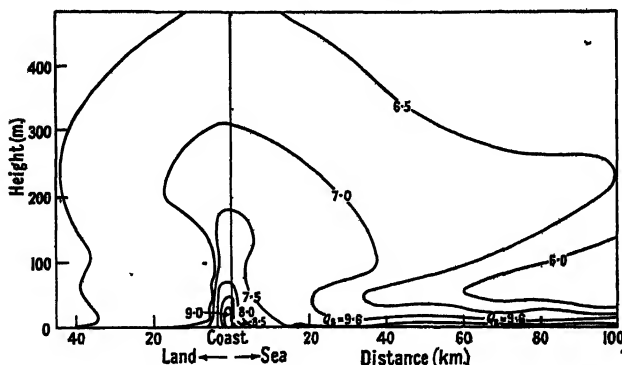


Figure 18. Isopleths of specific humidity (in gm/kg.), c. 1430 local time, 8th December 1946.

interference between two or more trapped modes within the duct. This latter possibility is unlikely in view of the fact that all four channels behave similarly, for it would be expected that different frequency or different height of source would change the position of such islands.

(v) *The Sea Breeze Complication in Summer*

In a light 'nor'west' situation, with a warm land surface, a sea breeze circulation at the coast frequently develops, and this may occur even in winter. A small-scale circulation is shown in Figures 17 and 18, the on-shore current of air being

about 100 metres deep at the coast, the air above passing off-shore. The sea breeze did not penetrate more than about 10 km. inland on this day, 8th December 1946, the sea air being rapidly warmed by the ground and ascending to pass out to sea again in the off-shore drift above 100 to 120 m. The isopleths show the circulation extending about 15 kilometres seawards, but relatively moist air is carried above 100 m. out to more than 100 km., though the shape of the isopleths above 100 m. between 10 and 60 km. off-shore is open to question, as observations in that region were scanty.

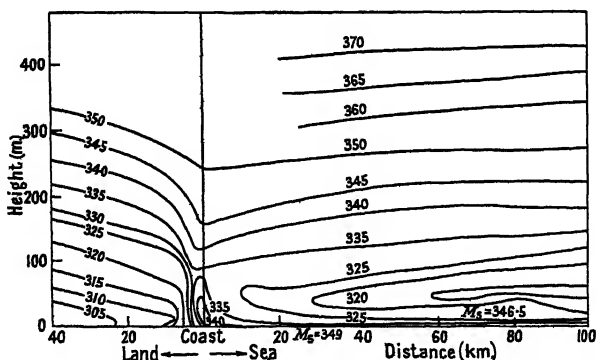


Figure 19. Isopleths * of modified refractive index, M , c. 1430 local time, 8th December 1946.

* The values assigned to the isopleths over the sea '325, 320, 325' should read '330, 325, 330'.

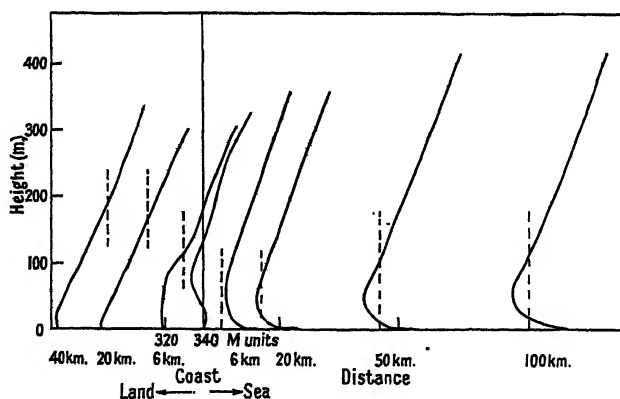


Figure 20. (M, h) curves, c. 1430 local time, 8th December 1946, ---- $M=330$.

The chief feature as regards the distribution of modified refractive index (Figures 19 and 20) is the existence of an elevated duct at the beach, and a small duct, or very steep slope of the (M, h) curve over a limited height interval, to a variable distance inland depending on the penetration of the sea breeze. The effect on the duct over sea is negligible beyond about 20 km. off-shore; at less than this range the effect of the sea breeze is to increase the height of the duct until it reaches a maximum at the beach (in this case 80 m.), after which it rapidly loses height and intensity as we pass inland. The net effect on radiation propagated at right angles to the coastline is practically as though there were no sea breeze. However, radiation propagated along the coast could be markedly affected by the presence of the elevated inversion.

(vi) *Some peculiar Situations*

The remaining situations have been selected because of some peculiarity in the meteorological situation which has a pronounced effect on propagation.

(a) *Low duct.*

On 23rd October 1947, a strong (force 6 to force 8) and gusty 'nor'wester' was blowing off-shore throughout the day. Although the temperature excess, $16-22^{\circ}\text{F.}$, and humidity deficit, varying between 2 and 4.5 gm/kg. , were ample for a strong duct to form off-shore, there was apparently so much turbulence near the surface of the sea that sufficiently sharp gradients were not maintained to give an inversion in the (M, h) curve above a height of about 15 m. The result was an intense low level duct up to a maximum height of 18 m. out to the full range covered, 160 km. from the coast (Figure 21). The relative effect on

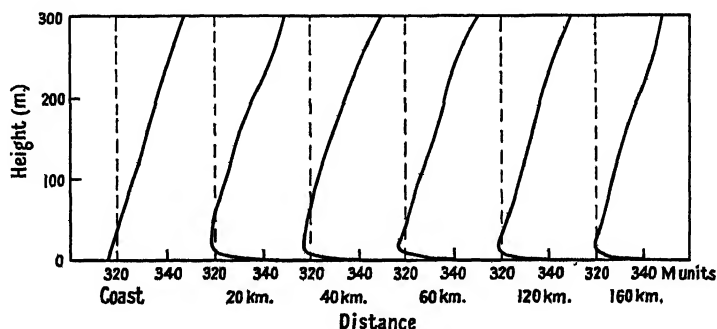


Figure 21. (M, h) curves at various ranges off-shore, 23rd October 1947, ---- $M=320$.

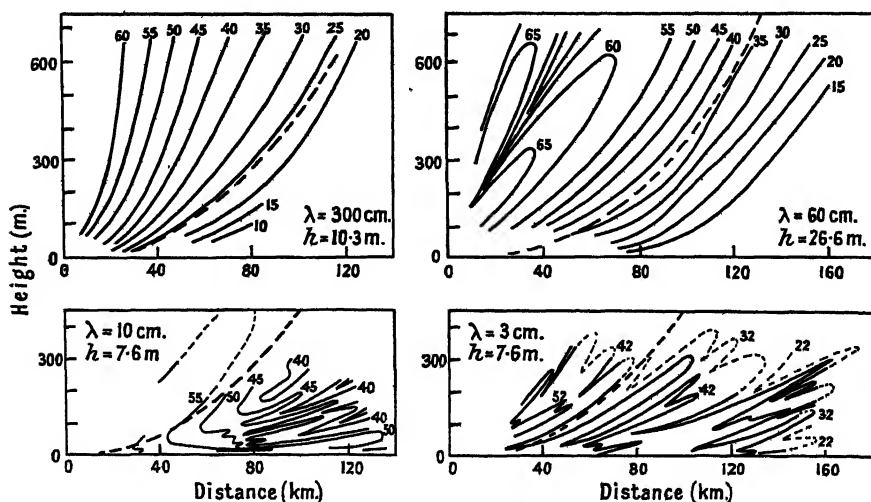


Figure 22. Field strength contours, 23rd October 1947, ---- horizon line for a standard atmosphere.

propagation at the different wavelengths is striking (Figure 22). At 300 cm. the effect is negligible, and at 60 cm. only very slight. Both lower sited microwave channels are effectively trapped, though the trapping at 3 cm. is not as great as would be expected when compared with 10 cm. This is possibly due to the greater scattering of the shorter wavelength energy at the rough sea surface.

at which range it weakened considerably, due largely to the change in surface temperature mentioned above. Beyond 120 km. there was considerable subsidence of warm dry air, and this caused the duct to increase in strength once more

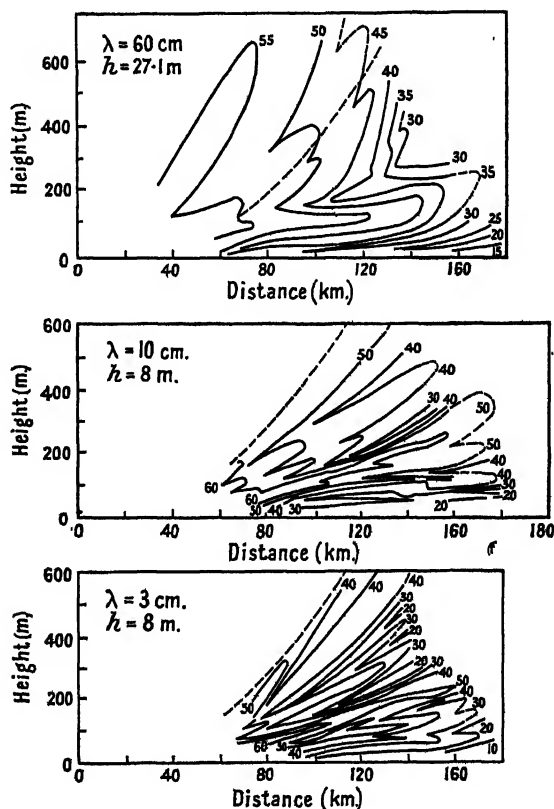


Figure 24. Field strength contours, c. 16 hours, 4th September 1947, --- horizon line for a standard atmosphere.

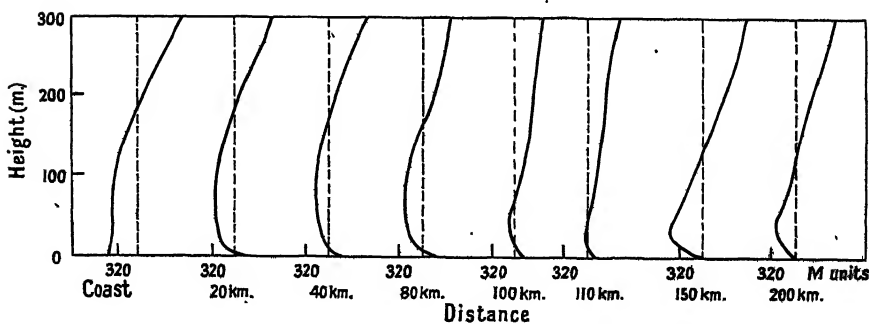


Figure 25. (M, h) curves at various ranges off-shore, 6th October 1947, --- $M=330$.

(Figure 25). The effect of this refractive index structure is clearly shown by the contours of field strength (Figure 26). On all three wavelengths shown, 60 cm., 10 cm., 3 cm., there is considerable trapping of energy in the lower levels out to a

range of 100 km., after which leakage is very pronounced as the duct weakens and the refractive index gradient diverts the energy upwards. Beyond 120 km., as the duct becomes stronger again, the contours indicate renewed trapping of energy close to the surface. The effect is least on the 3 cm. wavelength as a much greater proportion of energy is trapped by the intense part of the duct which is always present immediately above the surface.

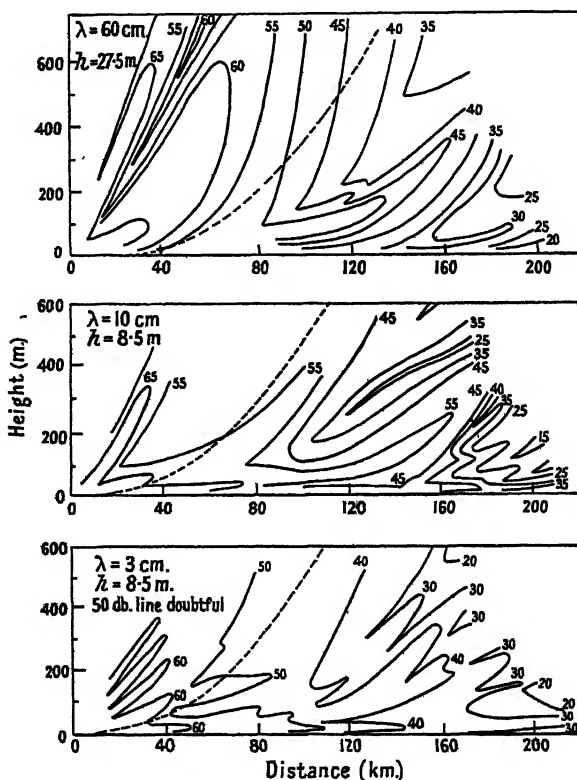


Figure 26. Field strength contours, 6th October 1947,
 --- horizon line for a standard atmosphere.

§ 4. CONCLUSIONS

A survey of the data that have been obtained allows us to draw some general conclusions concerning the phenomenon of duct formation and super-refraction produced by Föhn winds over the Canterbury Bight during the day-time. These conclusions are drawn from inspection of the data only and are therefore largely qualitative; no statistical or theoretical treatment has hitherto been attempted in New Zealand, although a certain amount has been carried out at the Telecommunications Research Establishment (Macfarlane 1948).

The Canterbury 'north-wester' has a temperature excess over the sea surface temperature of 10 to 20° F. in winter and up to 30° F. in summer: it has a humidity deficit relative to saturated air at the sea temperature of between 1 and 6 gm/kg. The resulting M deficit is usually between 15 and 30 Munits and an M inversion of this magnitude is formed immediately off the coast. The height of the inversion, i.e. the duct height, increases rapidly to about 30 m. in the first 10 or 20 km.

from the coast where the Föhn air characteristics are quickly modified by the relatively cool sea. The duct then rises more slowly to between 90 and 130 m. at a range of 60 to 120 km. off-shore as eddy diffusion carries the effects of the surface modification to greater heights. Associated with the increase in height of the duct is a slight decrease in intensity since the M inversion takes place at a higher value of M . Beyond 120 km. the duct tends to weaken and diminish in height slowly but this tendency is complicated by variations in sea temperature changing the surface value of M . At ranges of 160 to 220 km. the (M, h) curve is characterized by an inversion of about 5 Munits at the surface (up to 10 or 15 m.) and a nearly constant value above up to about 150 m.

The height of the duct in the first 60 km. off-shore appears to have a rough dependence on wind velocity, the greater the wind speed the lower the duct at a given range. This effect is not very clearly marked and does not extend over very wide limits, and for practical purposes can probably be neglected except when turbulence is excessive. In winds of strength above force 6 turbulence is usually so great that steep gradients of temperature and humidity are not maintained over large height intervals and the resulting duct, although intense, remains low (<20 m.) out to at least 160 km. from the coast.

The effect of the ducts usually encountered in a 'north-west' situation in daytime on radio propagation at various wavelengths is as follows:

- (i) Radiation on wavelengths of 10 cm. and 3 cm. is strongly trapped in the duct to give a maximum signal at a height of 15 to 30 m. of strength ± 10 db. relative to free-space signal level up to ranges of 200 km. from the coast.
- (ii) Radiation on a wavelength of 60 cm. is usually trapped to give a maximum signal at heights between 30 m. and 90 m. This signal is usually 5 to 10 db. below free-space level at a range of 100 km. and 15 to 20 db. below at a range of 200 km. There is no trapping on this wavelength when the duct height is less than about 50 m.
- (iii) Radiation on a wavelength of 300 cm. suffers only partial trapping of the first propagation mode and no peak in signal strength is found in the duct.

Field strength in the diffraction region below the horizon is considerably increased by leakage from the duct. For 10 cm. and 3 cm. the field pattern in this region has a very complicated structure but seldom results in a signal strength as great as that found in the duct.

The duct has negligible effect on propagation at angles of elevation exceeding $\frac{1}{4}^\circ$.

The presence of a local sea breeze beneath the off-shore drift increases the height of the duct locally but has very little effect on propagation at right angles to the coastline.

To sum up, the degree of trapping on microwave frequencies (3,000 Mc/s. and above) appears to vary only a little over quite a wide range of conditions. The practical result of this is that an approximate quantitative forecast of super-refraction conditions at these frequencies in a 'north-westerly' situation in Canterbury, or in other parts of the world where similar weather situations exist, may not be very difficult. Full analysis, however, especially of the meteorological data, is a task of considerable complexity, and will not be completed for a year or two.

ACKNOWLEDGMENTS

This investigation was carried out as part of the research programme of the Departments of Scientific and Industrial Research, United Kingdom and New Zealand, by whose permission this account is published.

The authors wish to acknowledge the contributions of the many members of these Departments who took part in the field observations and in the preliminary analysis work in New Zealand, and in particular the contribution of Dr. F. E. S. Alexander of Raffles College, Singapore, formerly with the Radio Development Laboratory of the Department of Scientific and Industrial Research, New Zealand, who was responsible for the initial organization of the project.

REFERENCES

- BOOKER, H. G., and WALKINSHAW, W., 1946, *Meteorological Factors in Radio Wave Propagation* (London: The Physical Society), p. 80.
 ENGLUND, C. R., CRAWFORD, A. B., and MUMFORD, W. W., 1935, *Bell Syst. Tech. J.*, **14**, 369.
 KATZIN, M., BAUCHMAN, R. W., and BINNIAN, W., 1947, *Proc. Inst. Radio Engrs., N.Y.*, **35**, 891.
 MACFARLANE, G. G., 1948, *Proc. Phys. Soc.*, **61**, 48.
 MEGAW, E. C. S., 1946, *J. Instn. Elect. Engrs.*, **93** (Pt. III A), 79.
 PRICE, W. L., 1948, *Proc. Phys. Soc.*, **61**, 59.
 SMITH-ROSE, R. L., and STICKLAND, A. C., 1946, *Meteorological Factors in Radio Wave Propagation* (London: The Physical Society), p. 18.
 Report of the Physical Society and Royal Meteorological Society Conference, 1946, *Meteorological Factors in Radio Wave Propagation* (London: The Physical Society).
 Summary Technical Report of the Committee on Propagation, N.D.R.C., 1946 (New York: The Academic Press Inc.).

Ionospheric Cross-Modulation: Techniques of Measurement

By C. C. NEWTON, F. J. HYDE AND H. G. FOSTER

Department of Electrical Engineering, The University of Birmingham

Communicated by L. G. H. Huxley; MS. received 20th March 1950

ABSTRACT. This paper describes the techniques developed at Birmingham University to measure the amplitude and phase of the ionospheric cross-modulation observed in the test transmissions which are described in detail elsewhere. The devices which were developed and the techniques of observation employed allowed trustworthy measurements to be made with comparative ease.

§ 1. INTRODUCTION

THIS paper describes in detail the chief techniques that were developed by a team of observers at Birmingham for the systematic investigation of ionospheric cross-modulation (radio wave interaction).

An account of the results of this investigation has already appeared elsewhere (Huxley 1950), but no systematic description of the experimental techniques employed has been published hitherto.

As described elsewhere (Huxley, Foster and Newton 1948, Ratcliffe and Shaw 1948, Huxley 1950) the technical problem is to measure the depth of transferred modulation T_o which is impressed at an audio frequency $\omega/2\pi$ on a

radio carrier wave (the wanted wave) during its passage through the ionosphere. The modulation T_ω is in fact transferred to the previously unmodulated wanted wave within a region of the ionosphere which is irradiated by a second radio wave (the disturbing wave) which is already modulated with the tone of frequency $\omega/2\pi$ when transmitted; that is to say, cross modulation takes place from the modulated disturbing wave to the unmodulated wanted carrier wave.

Since the modulation depths T_ω are small and in practice of the order of a few per cent, special techniques were required to measure them accurately.

It was also of importance, for reasons described in the references already given, to measure the phase of the modulation T_ω relative to the modulation M_ω of the same frequency on the direct ground wave received at the point of observation from the disturbing transmitter. The retardation in phase of T_ω relative to M_ω at the same point of observation is denoted in what follows by χ_ω .

§2. MEASUREMENT OF PERCENTAGE DEPTH OF TRANSFERRED MODULATION T_ω

Figure 1 is a block diagram of the system used to measure the modulation T_ω and the phase retardation χ_ω at any frequency between 50 c/s. and 2,000 c/s.

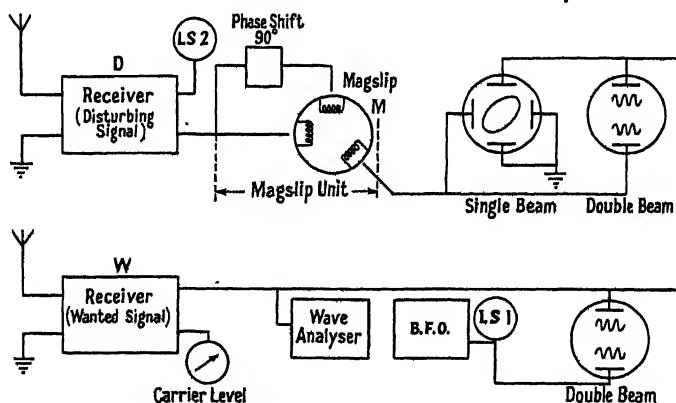


Figure 1. Block diagram of receiving system.

D and W are a pair of commercial communications receivers, tuned respectively to the disturbing and wanted waves. In each receiver the detector and audio output stages are modified. In D the detector load was changed in order that the modulation depth of 80% on the disturbing wave could be handled without distortion. The original detector and output stages of W were replaced by those shown (Figure 2) together with a built-in valve voltmeter to give the strength of the carrier of the wanted wave. The output stages in both D and W are cathode followers.

Consider the audio-frequency output from W. This is fed in parallel to a Marconi wave analyser (bandwidth 4 c/s.) and to one of the Y-terminals of a double-beam oscilloscope. To the other Y-terminal an audio-frequency voltage is applied from a beat-frequency oscillator (B.F.O., Figure 1) fitted with an output meter. Thus, with a linear time base, a pair of sinusoidal traces appears on the face of the tube.

The beat-frequency oscillator also excites a loudspeaker LS1 through an isolating amplifier, consequently the B.F.O. can be adjusted to give 'zero beat' with the tone from the loudspeaker LS2 driven by D. This is a convenient method of synchronizing the traces on the oscilloscope. The amplitudes of the two traces can be made identical by superimposing the traces and appropriately adjusting the output from the B.F.O.

The reading of the output meter of the B.F.O. provides, therefore, a relative measure of the audio signal output from W. Alternatively, the output from W may be measured directly by means of the wave analyser.

In order to obtain the absolute percentage value of the transposed modulation a preliminary overall calibration of the system is carried out by feeding into W a carrier modulated to known small percentages.

The factors by which it is necessary to multiply the B.F.O. readings or the wave analyser readings so that when they are divided by the readings of the carrier meter the absolute percentage modulations T_ω are obtained, were thus found for any modulation frequency $f = \omega/2\pi$. Under actual conditions of observation, with the normal strength of carrier, transferred modulations T_ω as small as 0.2% could be measured.

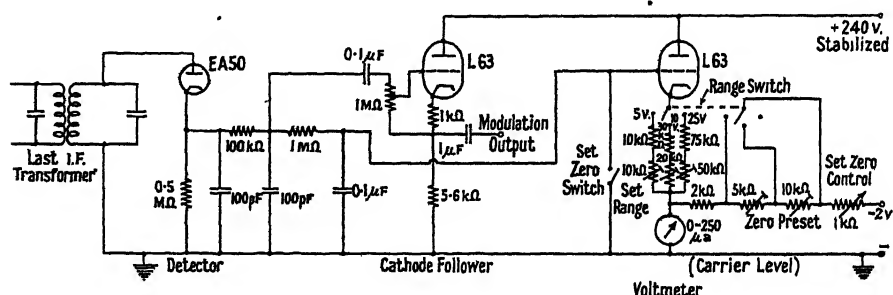


Figure 2. Detector and output units.

§ 3. MEASUREMENT OF RELATIVE PHASE RETARDATION

In order to determine the height of the mean centre of cross-modulation it is necessary to find the phase retardation of the transferred modulation at the receiver, relative to the modulation received on the ground wave from the disturbing transmitter.

The method used to measure this quantity at each frequency of $f = \omega/2\pi$ is indicated in Figure 1.

The output from the receiver D which is tuned to receive the disturbing wave excites the loudspeaker LS2, as mentioned above, but is also fed to the magflip unit. From the input voltage to this unit two alternating currents of equal amplitude but in phase quadrature are made to flow in the respective stators of a two-phase magflip. A rotating magnetic field of constant amplitude is thus excited within the magflip which induces in its rotor an electromotive force whose amplitude is independent of, but whose phase depends upon, the angular position of the rotor. The output voltage from the rotor is applied, through an amplifier, to one pair of deflecting plates of a single-beam oscilloscope. The transferred modulation T_ω from the receiver W is applied to the other pair of deflecting plates so that in general an elliptical trace appears on the face of the oscilloscope. However, by turning the rotor of the magflip the phase of the

The phase shift of 90° between the stator currents at all frequencies used was produced by means of a Miller integrator (V_8). Unfortunately, the amplitude of the output from the Miller integrator, with constant input, is inversely proportional to the frequency. To overcome this defect two procedures are obvious. A correcting network giving amplification proportional to the frequency with negligible phase shift must be introduced either before or after the integrator. In practice it was found preferable to place the network before the integrator, since in the alternative position serious hum was introduced.

The correcting network comprises valves V_1 to V_6 with their associated circuit components. Valves V_5 and V_6 are amplifiers with valves V_3 and V_4 in their cathode circuits to provide negative feedback controllable by the D.C. voltage output of the diode V_2 . The input to V_2 is developed across a circuit resonant at 40 c/s., which lies just below the lowest modulation frequency used in practice. The tuned circuit L, C together with the series resistance R form a potentiometer such that the input to V_2 diminishes as the frequency increases. The negative control voltage on the grids of valves V_3 and V_4 thus decreases as the frequency rises, with consequent reduction in their A.C. resistances. Thus the gain of stages V_5 and V_6 increases in proportion to the frequency up to 1,000 c/s., after which there is slight departure from proportionality. It follows, therefore, that the signal, when it reaches the input to the integrator from the cathode follower V_7 , has been amplified by a factor proportional to the signal frequency. There are now two voltages, the original signal applied to V_{12} and the voltage in quadrature with it which drives V_9 . The valves V_9 and V_{12} are cathode followers which feed the primary windings of the identical step-up transformers T_1 and T_3 . In order that the currents shall be the same in the primary windings of the transformers the input to the grid of V_9 is controlled by a potentiometer. The secondary voltage from T_1 drives the pair of cathode followers V_{10} and V_{11} in push-pull. These cathode followers are used to drive a current in the primary of the output step-down transformer T_2 whose secondary is coupled to one of the stators of the magflip. The other stator is driven from an identical circuit following V_{12} .

A rotating field is thus produced in the magflip which is effectively circular over the whole range of input frequencies employed.

§ 5. UNIFIED DISPLAY OF AMPLITUDE AND PHASE

With the system shown in Figure 1, amplitude and phase are found independently and thus require the presence of at least two observers. It was thought to be of interest and use to develop the system so that the amplitude-phase information is given on the face of a single oscilloscope in the form of an (r, θ) display. This object was achieved successfully by means of the system shown in Figure 4. This system is a replica of that given in Figure 1 but with the following additional units: a magflip M_2 and a switch unit S ganged to the original magflip M_1 , a rectifier unit R and a single-beam cathode-ray oscilloscope. The transferred modulation T_ω from the receiver W is fed into the rotor of M_2 and induces voltages in its stators whose amplitudes are proportional to $\cos \theta$ and $\sin \theta$, where θ is the angular position of the rotor. These voltages are rectified and applied to the respective pairs of deflecting electrodes of the single beam oscilloscope, and produce a radial deflection of the spot proportional to T_ω in the direction θ . However, in order to obtain radial deflections for all directions other than those in a single quadrant it is necessary to introduce a switch unit to ensure that the

The amplifier, consisting of valves V_{15} and V_{19} and associated components, which is used to feed the output of receiver W into the rotor of magslip M_2 , is similar in general principles to those circuits previously discussed in relation to magslip M_1 . However, because of the reduced response at low frequencies of the various transformers used, it was found necessary to introduce a correcting network in order to maintain linearity between the input signal applied to the system and the subsequent D.C. voltage output from the rectifier unit. This network is shown in the anode circuit of valve V_{16} .

The output from the first stator of magslip M_2 is amplified in the circuit containing valves V_{20} to V_{23} , and applied by means of transformer T_5 to a bi-phase rectifier V_{24} , the output of which is smoothed in normal manner by the low-pass filter shown. The resulting D.C. voltage is developed across the 250 k Ω potentiometer, which is adjusted in practice to give an adequate deflection of the spot on the face of the cathode-ray tube, when the signal output from the receiver W is at the lowest frequency used in the tests.

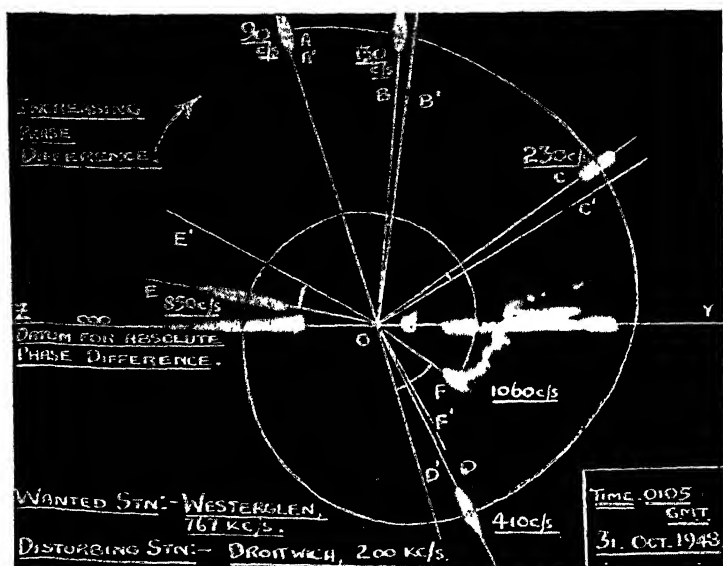


Figure 6. Photograph of (r, θ) display of ionospheric cross-modulation.

In this way, the amplitude of the D.C. voltage output of the system is proportional to the amplitude of the transferred modulation T_w accepted by the receiver W, at any fixed angular setting θ of the rotor of magslip M_2 .

The output from the second stator of magslip M_2 is fed to an identical set of circuits, and thus we have two D.C. signals whose amplitudes are each proportional to the transferred modulation T_w , and individually to $\cos \theta$ and $\sin \theta$ respectively.

However, since the polarity of these D.C. voltages is positive with respect to earth, and is independent of θ , a switch unit in the form of a simple commutator is ganged to the rotors of magslips M_1 and M_2 . One D.C. voltage, passing through one half of the switch unit, is applied to the X deflector plates of the cathode-ray

tube and the other voltage similarly to the Y deflector plates. The ganging of the switch unit to the rotors of magslips M_1 and M_2 is arranged in such a way that the spot on the face of the display tube appears in the same quadrant as the angular setting of the rotors.

The performance of the system may be judged from Figure 6 which is a photographic record of an actual test transmission. In addition to the actual amplitude-phase polar record, the phase error is corrected at each frequency and the true phase angle is indicated in each case by a radial line.

It was found in practice that the straightforward technique described in §§ 2 and 3 was simpler to use and more flexible.

ACKNOWLEDGMENTS

The authors have pleasure in acknowledging their appreciation of the invaluable assistance given by the Engineering Research Department of the B.B.C. in providing transmission and to Messrs. H. L. Kirke and R. A. Rowden, the observers at Tatsfield and Caversham and the engineers at the transmitters in particular.

They are also indebted to Professor A. Tustin for encouragement.

They finally wish to record their appreciation of the general guidance throughout the work given by Dr. L. G. H. Huxley who suggested some of the techniques which the authors were able to develop.

REFERENCES

- HUXLEY, L. G. H., 1950, *Proc. Roy. Soc. A*, **200**, 486.
HUXLEY, L. G. H., FOSTER, H. G., and NEWTON, C. C., 1948, *Proc. Phys. Soc.*, **61**, 134.
RATCLIFFE, J. A., and SHAW, I. J., 1948, *Proc. Roy. Soc. A*, **193**, 311.
-

LETTERS TO THE EDITOR

The Measurement of Microwave Power at Wavelengths of 3 cm. and 10 cm.

Collard, Nicoll and Lines (1950) have recently pointed out that large discrepancies may occur in the measurement of microwave power at wavelengths below 3 cm. using thermistor milliwattmeters. In this note we wish to give a preliminary report of microwave power measurements recently carried out here on the absolute calibration of milliwattmeters for operation in the 3 cm. and 10 cm. wave bands.

The apparatus used in the measurements for both wavebands were similar in principle and consisted of a generator, either a pulse magnetron or a klystron (both capable of delivering mean powers of the order tens of watts into a matched load), connected to a waveguide run. The latter was terminated in a matched wedge constant flow calorimeter, and a waveguide Bethe single-hole directional coupler was inserted between the generator and calorimeter. Thus, if the power entering the calorimeter is determined and the fraction of the power abstracted by the directional coupler (coupling factor) is known, the absolute power delivered into a milliwattmeter matched to the low power guide of the directional coupler may be calculated. The coupling factors of the directional couplers were about 44 db., as determined by a piston attenuator of accurately known characteristics, and hence for a dissipation of 25 watts in the calorimeter the power delivered to the milliwattmeter was about 1 milliwatt in each case.

The milliwattmeters used in the experiments were :

- (i) *10 cm. band.* A thermistor head designed by R. C. Robbins at R.R.D.E.; a short description of this instrument has already been published (Street 1949).

(ii) 3 cm. band. A.S.R.E. thermistors, with ceramic protective tubes, incorporated into a T.R.E. design of waveguide thermistor mount.

(iii) 3 cm. band. Sperry type 821 barretter capsule in a type 82X mount (Montgomery 1947).

All the instruments were standardized by D.C., and the results of calibration are given in the following table:

Milliwattmeter type	(i)	(ii)	(iii)
Ratio of $\frac{\text{Absolute power}}{\text{Indicated power}}$	1.04	1.10	1.04*

* This result was obtained by comparing the power indications of instruments (ii) and (iii) at a level of 1 milliwatt of c.w. power in order to avoid the possibility of burn-out of the barretter by peak pulse power.

It is estimated that the accuracy of this method of calibration is within approximately 2 per cent. Further experiments showed that the reproducibility of milliwattmeter readings was 2-3 per cent for type (ii) milliwattmeters when different samples of thermistor were incorporated in the same mount. From these measurements it appears that there is no significant difference between the absolute power input and the powers indicated by milliwattmeters (i) and (iii). However, the readings of milliwattmeters of type (ii) may be in error by about -10 per cent. Further details of this work will be published at a later date.

Acknowledgment is made to the Chief Scientist, Ministry of Supply, for permission to publish. Thanks are also due to the Chief Superintendent, Radar Research and Development Establishment, for the provision of essential apparatus.

The University,
Nottingham,
1st June 1950.

R. STREET.
P. D. WHITAKER.

COLLARD, J., NICOLL, G. R., and LINES, A. W., 1950, *Proc. Phys. Soc. B*, **63**, 215.

MONTGOMERY, C. G., 1947, *Technique of Microwave Measurements* (New York and London: McGraw-Hill), Chap. 3.

STREET, R., 1949, *Proc. Inst. Elect. Engrs.*, **96**, Pt. II, 391.

Photoconductive Cells of Cadmium Selenide

Polycrystalline layers of CdSe have been produced by methods previously described (Schwarz 1948). These cells exhibit an appreciable sensitivity for electromagnetic radiation within a range starting from the x-ray region, covering the Schumann, ultra-violet and visible region and extending into the near infra-red. The spectral distribution of sensitivity at room temperature for the visible, near ultra-violet and infra-red is given in

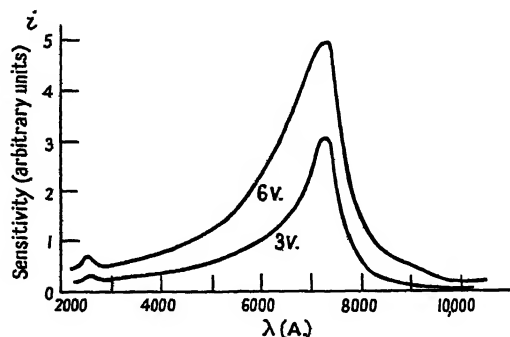


Figure 1.

Figure 1 for different voltages across a cell with a sensitive area of 0.2 mm. × 10 mm. There is a sharp drop of sensitivity after the maximum at approximately 0.72 μ. The position of the maximum varies somewhat for different cells.

The sensitivity of a cell with a sensitive area of 10 mm. \times 0.2 mm. to the radiation of a tungsten filament lamp operated at a colour temperature of 2,850° K. has been measured by the National Physical Laboratory with the following results :

Illumination (lumens/m ²)	2.15	21.5	215
P.D. across cell (volts)	12	12	12
Cell current (μ a.)	21.6	365	2800
Cell current per lumen (amp/lumen)	5.0	8.5	6.5

The dark current of the cell for a potential difference of 12 volts was less than 0.2 μ a. The dark resistance of the cells can be varied within very wide limits and may be as high as 10^{10} ohms at room temperature.

The response of a cell to x-rays from a Cu anode for different electron currents I , voltages E and cell distances from the anode d are shown in Figures 2 and 3. The photocurrent i in microamps was measured for a field of 1,100 v/cm. across the cell.

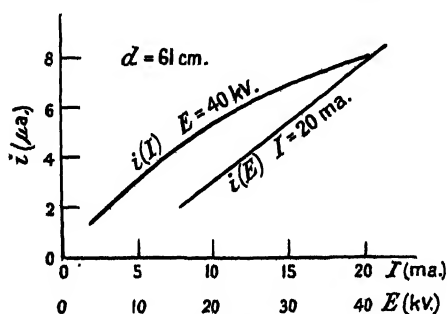


Figure 2.

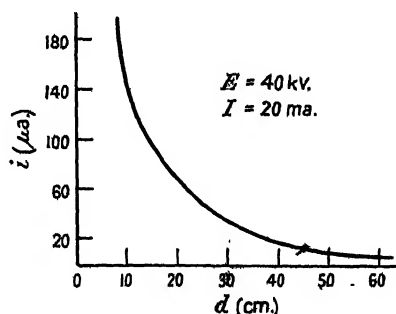


Figure 3.

Preliminary speed measurements indicate that cells with extremely high D.C. sensitivities are relatively slow, as would be expected. At a chopping frequency of 200 c/s. for the light from a tungsten lamp the output may drop to one third or one fourth of the D.C. value, but there seems to be a strong dependence of the speed on light intensity and applied voltage. As, on the other hand, it has been possible to minimize slow, secondary effects (fatigue and creep) to a very small fraction of the photocurrent, these cells should prove of great practical importance for D.C. measurements.

Contrary to the method for the production of photoconductive single crystals of CdS and CdSe (Frerichs 1946) the methods used here involve the application of oxygen and it has been found that all the properties of the cell are governed by the amount and form of oxygen present in the layer. It has been suggested (Schwarz 1949) that adsorption of oxygen on the grain boundaries is one essential condition for the production of photocells with a high quantum yield and the experiments on CdSe strongly support this view. This matter will be discussed in detail elsewhere.

I would like to thank Mr. C. G. Cannon, University of Reading, for help with measurements in the Schumann region, Mr. J. H. Pratt, Services Electronics Research Laboratories, Baldock, for speed measurements, and Hilger and Watts Ltd. for permission to publish this letter.

Hilger and Watts Ltd.
98 St. Pancras Way,
Camden Town, N.W. 1.
8th June 1950.

E. SCHWARZ.

FRERICHS, R., 1946, *Naturwissenschaften*, **33**, 281.

SCHWARZ, E., 1948, *Nature, Lond.*, **162**, 614; 1949, *Proc. Phys. Soc. A*, **62**, 530.

REVIEWS OF BOOKS

Gas Tables, by JOSEPH H. KEENAN and JOSEPH KAYE. Pp. x+238. 2nd Edition. (New York: John Wiley and Sons, Inc.; London: Chapman and Hall, Ltd., 1948). 30s. (Published in 1945 under the title *Thermodynamic Properties of Air, including Polytropic Functions*.)

These tables are intended to do for the engineer using air or similar gases what the Steam Tables do for the steam engineer. By assuming air, or the products of combustion, to be gases which obey the law $pV=RT$, the tables are greatly simplified, for the enthalpy and internal energy are then functions of one variable only, and the ratio of volumes at two temperatures along an adiabatic is the same whatever adiabatic is selected. Instead, therefore, of a table of double entry as used in steam engineering, a table of single entry can give volume ratio and the fundamental thermal properties. The error so introduced is discussed on pages 203 to 204.

The tables are a revision of *Thermodynamic Properties of Air* by the same authors. The thermal properties are based on spectroscopic data, using absolute zero as the zero for energies, and adopting the latest values for the gas constant, the velocity of light and the value of the international joule in absolute units.

The tables, 64 in number, are designed to cover the conditions and to give directly the quantities needed for turbine calculations and for jet design work. They include a table for air, others for products with various quantities of excess air, for hydrogen, hydrocarbons, carbon monoxide and argon. Other tables deal with conditions along a Rayleigh line or a Fanno line, with different values of the polytropic index, and yet others are purely mathematical tables ($\log N$, and fractional powers). The whole collection should be of great value to technologists in this field.

J. H. A.

A Manual of Vacuum Practice, by L. H. MARTIN and R. D. HILL. Pp. 82+34, Appendix. 1st Edition. (Melbourne: University Press, 1947; British Agents: Cambridge University Press.) 10s. 6d.

This little volume is probably the best of the manuals and handbooks on vacuum technique which have been offered to us. It omits the detailed discussion of kinetic theory principles which can be found in the standard textbooks on the subject and presents them in a form immediately useful for the calculation of vacuum systems. The theory underlying the various types of pump and gauge is sketched and profuse references to papers and textbooks are given (95 in all).

In chapter 1, the kinetic theory formulae relating to molecular flow etc. are given, with some examples of how to calculate the types of pump, sizes of piping etc. required in actual vacuum systems, including consideration of what happens when a burst of gas is released. Chapter 2 describes the various types of pressure gauge and includes a note on the difficulties and errors in the use of the ionization gauge which have recently come into prominence. Chapter 3 is devoted to pumps and the characteristics of a number of modern rotary and diffusion pumps are illustrated. The importance of keeping water out of oil diffusion pumps—an important practical point—is mentioned. The final chapter is on vacuum plumbing; various types of rubber-sealed valves are described, also the Wilson seal for transmitting motion into a vacuum system, dimensions for various shaft sizes being given. Leak finding is discussed, including the use of the palladium ionization gauge.

The 18 appendices contain information about the properties of various substances used in vacuum work, including getters. There are also descriptions of the manufacture of copper-to-glass sales and wire seals through glass and silica.

No misprints of any importance were detected in the book, which can be recommended to students of physics and those engaged in the "vacuum consuming" industry.

M. D. ARMITAGE,

Tables of Sines and Cosines to Fifteen Decimal Places at Hundredths of a Degree, NATIONAL BUREAU OF STANDARDS, APPLIED MATHEMATICS LABORATORY. Pp. viii + 95. (Washington: U.S. Government Printing Office, 1949.) \$0.40.

Of the three systems for measuring angles (the radian, the degree with sexagesimal sub-division and the degree with decimal subdivision), sometimes one and sometimes another is the favourite of the table maker. For many years, when 8 or less figures were wanted, the degree, minute and second were much better catered for than were the others, though now they are all properly tabulated at convenient intervals in easily accessible volumes.

When we turn to many-figure tables, however, the position is somewhat different, and it appears that there has been no table in degrees and decimals for more than 10 figures since the pioneer work of Briggs, in 1633. He gave 15 decimals, at intervals of 0.01° , and his work has now been repeated by the Mathematical Tables Project. They have recalculated the values at each degree, using 25 decimals, have sub-tabulated to obtain the values at 0.01° and then sub-tabulated these. This gave a draft to 18 decimal places, which was checked by differencing, and then rounded off and checked again by differencing. It would appear that the volume is printed from this typescript by a photographic method, thus avoiding the possibility of errors by compositors. There is no suggestion that the final table has been read against Briggs's table, in which the only known error is at 16.49° .

Linear interpolation in these tables can be relied on everywhere for 8 decimals, and second differences suffice for full accuracy. These differences are tabulated opposite every entry, and the coefficients required in Everett's formula are given in a supplementary table. Inverse interpolation can be effected either by taking $\arcsin x$ from tables of that quantity, and then converting to degrees, or by any of three methods explained in the Introduction to the present tables.

There is a supplementary table of $\sin x$ to 30 decimals, at intervals of 1° . This was computed by Herrmann, who gave also a table of tangents; for the present volume, his values have been checked by the relation $\cos x \tan x = \sin x$ and by the formula

$$\sin 1^\circ(1 + 2\sum_1^{89} \sin x^\circ) = 1 + \cos 1^\circ.$$

The tables are bound in stiff paper, and owners will probably think it worth while to have them properly bound.

J. H. A.

Surface Tension and the Spreading of Liquids, by R. S. BURDON. Pp. xiv + 92. Cambridge Monographs on Physics. 2nd Edition. (Cambridge: University Press, 1949.) 12s. 6d.

The first appearance of Dr. Burdon's book on Surface Tension in 1940, as one volume of the series of Cambridge Physical Tracts, was a healthy reminder that in one branch at least of classical physics considerable progress is still taking place, and that the spirit of novelty is far from dead. In the original monograph we were reminded of many points of interest which are frequently glossed over in the standard treatises on the general properties of matter; and it must have been with considerable surprise that many readers learnt of the scantiness of our exact knowledge of such an apparently straightforward matter as the surface tension of mercury. All this was very much to the good, as it necessarily stimulated interest in a branch of physics which is an admirable subject of research for many of those to whom the elaborate and expensive apparatus associated with some other branches of physics is quite prohibitive.

Now, in the second edition, we have a complete revision of the original book, attractively presented in a cloth binding, albeit at a somewhat increased price. The chapter on the measurement of surface tension has been expanded to include some account of the measurement of small differences in tension. Greater attention is given in the new edition to the effects of adsorption and surface contamination; and progress is noted in the exact measurement of the surface tension of mercury. On the subject of the spreading of liquids on solid and liquid surfaces, Dr. Burdon's account, which forms a large part of the book, is the most

comprehensive up-to-date treatment which is available to the English reader. This section, like others, has been fully revised and now includes a short section on the effect of surface films in inhibiting evaporation from water. The section on angles of contact has been expanded to include reference to the newer methods of measurement, and Bate's observation of a 90° angle of contact between glass and carefully prepared mercury is recorded in the appropriate section. It is perhaps only right to say, however, that our knowledge of the angles of contact of liquid metals is still very scanty, and that real progress is possible only when the less subjective methods of observation are applied under conditions in which the degree of surface contamination is properly controlled.

Altogether, Dr. Burdon's book is a valuable contribution to the literature of what is still, unfortunately, known as classical physics, and in its new edition should be of even greater value both to the University student and to the industrial physicist who is concerned with the practical solution of the problem of wetting.

G. D. YARNOLD.

Elements of Aerodynamics of Supersonic Flows, by A. FERRI. Pp. x + 434. 1st Edition. (New York: The Macmillan Co., 1949.) 50s.

The aerodynamic theory of supersonic flow is receiving so much attention at the present time and being developed so rapidly that any book purporting to be up to date when written is inevitably no longer so by the time it gets published. Through no fault of the author this is to some extent true in this particular case as, during the past two years, many important advances have been made which are not even mentioned in the book. In this connection readers in this country will note with some surprise the unfortunate lack of references to recent British work. However, the book as a whole is a valuable addition to the literature on the aerodynamics of supersonic flows and makes a useful introduction to the subject. It discusses in detail the fundamental principles involved and the assumptions made in the development of the theory. The various problems considered are treated in a manner that is easily understood, and the material is presented clearly and concisely. In the main the book is written from the engineer's point of view rather than that of the mathematician, and it should be of great value to aeronautical engineers and designers who are concerned mainly with practical applications of the theory of supersonic flow.

Chapter 1 considers in general the basic equations of the flow and the physical nature of the phenomena associated with bodies moving faster than sound. In the following four chapters the theory of two-dimensional flow for small disturbances, the theory of shock waves and the theory of characteristics for rotational as well as irrotational flow are discussed.

Chapter 6 deals with the techniques used for measuring the physical quantities associated with supersonic flow, and a brief account is given of the principles involved in the use of the interferometric, the shadowgraph, and the schlieren methods for investigating density variations.

Various applications of the theories developed in the earlier chapters are given in Chapter 7. Pressure distributions on supersonic profiles of different shapes are determined and comparisons are made between theoretical and experimental results. The influence of viscosity, which is neglected in the theory, and effects due to the presence of boundary layers are also discussed. This chapter concludes with an account of the Buseman biplane and the basic ideas of its design.

The problems which arise in the design of effusers and diffusers are discussed in some detail in Chapters 8 and 9. Numerical examples are included to illustrate the methods of design described.

After giving an outline of the theory of supersonic flow in three dimensions in Chapter 10, the author proceeds to discuss in the next three chapters problems connected mainly with axially symmetric flow and particularly those associated with flow past bodies of revolution. Numerical examples are given of some of the cases considered, and these should form a useful guide to the application of the theory to the calculation of actual flows.

The last two chapters deal with the determination of the pressure distributions and aerodynamic forces on wings of various plan forms.

Tables which are useful in supersonic flow investigations are included at the end of the book, the value $\gamma = 1.40$ being assumed.

W. P. JONES.

CONTENTS FOR SECTION A

	PAGE
Prof. G. I. FINCH. The Sliding Surface (34th Guthrie Lecture)	785
Dr. P. TORKINGTON. The Planar Vibrations of Tetra-chloro-ethylene: an Example of a Complete Normal Coordinate Analysis	804
Miss A. J. BUZEMAN. The Electronic Structure and the Bond Lengths of Ovalene.	827
Dr. H. KUHN and Mr. G. K. WOODGATE. Hyperfine Structure and Nuclear Spin of Yttrium, ^{89}Y	830
Dr. J. R. HOLT and Mr. C. T. YOUNG. The Angular Distribution of Protons from the Reaction $^{27}\text{Al}(\text{d}, \text{p})^{28}\text{Al}$	833
Mr. R. W. PARSONS and Mr. C. H. COLLIE. Photodisintegration of the Heavy Elements	839
Dr. S. C. CURRAN, Mr. A. L. COCKROFT and Mr. G. M. INSCH. Gamma and Beta Ray Spectroscopy with Proportional Counters in Magnetic Fields	845
Dr. H. LAWTON and Dr. K. H. STEWART. Magnetization Curves for Polycrystalline Ferromagnetics	848
Mr. D. A. RICHARDS. Magnetite Inclusions in Mica	852
Dr. V. APPAPILLAI and Prof. A. W. MAILVAGANAM. Transition Effect of Extensive and Local Penetrating Cosmic-Ray Showers in Colombo	856
Mr. ALLADI RAMAKRISHNAN. A Note on the Size-Frequency Distribution of Penetrating Showers	861
Mr. D. BROADBENT, Dr. E. W. KELLERMANN and Mr. M. A. HAKEEM. The Density Spectrum and Structure of Extensive Cosmic-Ray Air Showers at Sea Level	864
Mr. S. L. MARTIN, Dr. E. H. S. BURHOP, Mr. C. B. ALCOCK and Dr. R. L. F. BOYD. The Scattering of Neutrons by Deuterons	884
Dr. F. C. BARKER. Schwinger Potential in Nuclear Forces	898
Mr. D. V. OSBORNE. The Rotation of Liquid Helium II	909
Letters to the Editor:	
Dr. M. BLACKMAN and Mr. V. F. G. TULL. On the Inner Potential of Metals.	913
Mr. L. E. BEGHIAN, Mr. M. A. GRACE and Dr. H. HALBAN. The $^{11}\text{B}(\alpha, \text{n})^{14}\text{N}$ Reaction	913
Mr. R. W. PARSONS, Mr. D. J. LEES and Mr. C. H. COLLIE. The Photo Threshold of ^{208}Pb	915
Mr. E. W. TITERTON. The Photodisintegrations $^6\text{Li}(\gamma\text{D})^4\text{He}$ and $^7\text{Li}(\gamma\text{T})^4\text{He}$	915
Prof. M. A. CATALÁN and Mr. R. VELASCO. Unequal g -Values for the Different Magnetic Levels of an Atomic Energy Level	917
Mr. H. D. RATHGEBER. The Semi-Diurnal Variations of Atmospheric Pressure and of Cosmic-Ray Intensity	920
Mr. C. D. MEE. The Mechanism of Colloid Agglomeration in the Formation of Bitter Patterns	922
Contents for Section B	923
Abstracts for Section B	923

ABSTRACTS FOR SECTION A

The Sliding Surface (34th Guthrie Lecture), by G. I. FINCH.

The Planar Vibrations of Tetra-chloro-ethylene: an Example of a Complete Normal Coordinate Analysis, by P. TORKINGTON.

ABSTRACT. A normal coordinate analysis of the planar vibrations of tetra-chloro-ethylene is carried out. For the B_{1g} , B_{2u} and B_{3u} modes the analysis is complete, including the calculation of the coefficients in the transformation to normal coordinates, the components in the displacements, the directions of the displacement vectors and the distribution of the potential energy, as well as the force constants, for a range of solutions; in each case, two alternative assignments are employed. For the A_g vibrations, several ranges of possible solutions are obtained. Curves are given in every case, and corrections for variable structural parameters calculated for selected solutions. The results for the various symmetry classes are discussed and correlated, and the most likely set of force constants in the valence-force potential function deduced. Details of the method of analysis are given.

The Electronic Structure and the Bond Lengths of Ovalene, by A. J. BUZEMAN.

ABSTRACT. The method of molecular orbitals has been applied to find the bond orders and the bond lengths of ovalene $C_{32}H_{14}$.

A comparison with the preliminary experimental values of Donaldson and Robertson shows general agreement.

Hyperfine Structure and Nuclear Spin of Yttrium, $^{89}_{39}Y$, by H. KUHN and G. K. WOODGATE.

ABSTRACT. By the use of a liquid-air-cooled hollow-cathode tube with very low current densities it was possible to resolve hyperfine structure in the arc spectrum of yttrium. The value of the nuclear spin was found to be $I=\frac{1}{2}$ and the sign of the magnetic moment to be negative. This agrees with recent results of Crawford and Olson, who determined the nuclear spin from the intensity ratio of a hyperfine-structure doublet in the second spark spectrum of yttrium.

The Angular Distribution of Protons from the Reaction $^{27}Al(d, p)^{28}Al$, by J. R. HOLT and C. T. YOUNG.

ABSTRACT. Experiments are described in which a differential ionization chamber was used in the measurement of angular distributions of disintegration particles. Results are given for seven groups of protons from the (d, p) reaction with aluminium in which deuterons having energies of 4.6, 5.8 and 7.5 mev. were used. The distribution curves are of complicated form, with a general tendency for the protons to be emitted in the forward direction.

Photodisintegration of the Heavy Elements, by R. W. PARSONS and C. H. COLLIE.

ABSTRACT. The photo thresholds for the emission of neutrons by the following elements have been determined. The threshold in mv. is given in the brackets following the symbol: Pt (6.1 ± 0.1); Au (8.1 ± 0.1); Hg (6.6 ± 0.2); Tl (7.3 ± 0.25); Pb (6.9 ± 0.1); Bi (7.2 ± 0.1); Th (6.0 ± 0.15); U (5.8 ± 0.15).

The neutrons were detected by the Szilard and Chalmers method of chemical concentration and the γ -rays obtained from a small betatron. Rough values of the total cross sections can be estimated from the results.

Gamma and Beta Ray Spectroscopy with Proportional Counters in Magnetic Fields, by S. C. CURRAN, A. L. COCKROFT and G. M. INSCH.

ABSTRACT. Proportional counters operating in a 'condensing' type of magnetic field can be used as β - or γ -ray spectrometers up to at least 2 mev. Results for P32 and annihilation γ -radiation are given.

Magnetization Curves for Polycrystalline Ferromagnetics, by H. LAWTON and K. H. STEWART.

ABSTRACT. Recently described developments have led to greatly improved agreement between theory and experiment for the magnetization curves of single crystals of iron and iron-silicon in the region where magnetization proceeds by rotation of domain vectors against the action of magnetocrystalline forces, i.e. in the range 10–500 oersteds. On the basis of the assumption that in polycrystalline material the magnetization is approximately uniform from grain to grain, which is rendered plausible by consideration of internal demagnetizing effects, a method is developed for deriving the magnetization curves of such material by averaging the fields required to produce a given magnetization in the different constituent grains, rather than by averaging the magnetizations produced in the different grains by a given field. The application of this method to specimens with known distributions of grain orientations is shown to give results in fair agreement with experiment.

Magnetite Inclusions in Mica, by D. A. RICHARDS.

ABSTRACT. The paper describes the investigation of the orientation of magnetite inclusions in mica using electron diffraction by transmission through composite films. The diffraction pattern obtained indicates that the magnetite is a single crystal with its (111) planes parallel to the cleavage plane of mica and one of the face diagonals parallel to the *b* axis of mica. The technique for preparing the thin films is described in some detail.

Transition Effect of Extensive and Local Penetrating Cosmic-Ray Showers in Colombo, by V. APPAPILLAI and A. W. MAILVAGANAM.

ABSTRACT. An account is given of the experiments on the transition effects of local and extensive penetrating cosmic-ray showers at Colombo (geomagnetic latitude 4° S.). A comparison of the present results with those reported for higher latitudes indicates that there is no observable latitude effect of the local penetrating showers. Contrary to the findings of Broadbent and Jánosy, the transition effect of the extensive showers in paraffin appears comparable in magnitude with the effect in lead.

A Note on the Size-Frequency Distribution of Penetrating Showers, by ALLADI RAMAKRISHNAN.

ABSTRACT. Heitler and Jánosy have calculated the probability that n mesons are produced due to nucleon-nucleon collisions by a fast nucleon in its passage through matter of a given thickness. In this note the results of Heitler and Jánosy are obtained by the direct method of writing the integro-differential equations of the problem using the Markoff nature of the process.

The Density Spectrum and Structure of Extensive Cosmic-Ray Air Showers at Sea Level, by D. BROADBENT, E. W. KELLERMANN and M. A. HAKEEM.

ABSTRACT. Showers have been examined in the density range 5–500 particles/m² using two trays of Geiger counters 5 m. apart, the discharges of individual counters being recorded using neon lamps. The observed showers are compared with a theoretical analysis based on the usual assumptions that the incident densities are uniform over the extent of the apparatus and distributed according to the spectrum $N(x) = Bx^{-\gamma}$ per hr. per m². The comparison shows (i) that in the density range examined the above spectrum applies only for a vanishingly small area of observation, with $\gamma = 1.425 \pm 0.022$, $B = 620$, and with no evidence for a change of γ with density, and (ii) that for the separation of 5 m. employed, the density can *not* be regarded as uniform over the extent of the apparatus. It is shown that this second conclusion is capable of explaining many of the differences found between the results of other authors obtained on the assumption of uniformity of density: in particular, it is concluded that there is yet no significant evidence for a change of γ with altitude.

The Scattering of Neutrons by Deuterons, by S. L. MARTIN, E. H. S. BURHOP, C. B. ALCOCK and R. L. F. BOYD.

ABSTRACT. The photographic plate method has been used to investigate the angular distribution of deuterons recoiling after neutron impact. The deuteron recoils were produced in a thin target of heavy wax placed in contact with a C2 Ilford nuclear emulsion.

For neutrons of energy 2.6–3.1 Mev., produced in the d-d reaction, 2,000 deuteron recoils were measured. Owing to the short range of the recoils produced when neutrons of low energy are scattered through small angles, the technique is only applicable to the study of the angular distribution in the range 90°–180° in the centre-of-mass system. For angles between 120° and 180° the results of the present work are in satisfactory agreement with those of previous experimenters using different methods, and they agree with the theoretical form of the distribution calculated by Buckingham and Massey on the assumption of ordinary forces between nucleons. This agreement is not decisive, however, since in this region the calculated distributions for ordinary and exchange forces do not lie far apart. For smaller scattering angles where a decisive test could have been made, the method is too inaccurate.

For neutrons of energy from 4.0 to 6.0 Mev. and from 6.0 to 9.0 Mev., obtained from a Ra-Be source, the angular distribution over a much larger range of scattering angle has been obtained from the measurement of 1,200 deuteron recoils. The angular distribution for neutrons scattered through angles near 180° appears to be rather steeper than for slower neutrons; but the ratio of the differential cross section at 90° to that at 180° is found to be about 0.35–0.40, of the same order as for neutrons in the lower energy range.

Measurements of 1,000 proton recoils from the d-d neutrons showed an angular distribution for neutron scattering angles between 90° and 180° that was constant within the limits of the statistical error.

Schwinger Potential in Nuclear Forces, by F. C. BARKER.

ABSTRACT. The properties of the potential derived from the mixed pseudoscalar-vector meson field are investigated and compared with the experimental data on the two-nucleon system. An attempt has been made to use methods which are more accurate than those formerly applied to this problem, and some of these may be useful in calculations with other potentials. It is found that agreement with experiment cannot be obtained for any choice of meson masses and coupling constants.

The Rotation of Liquid Helium II, by D. V. OSBORNE.

ABSTRACT. An experiment is described in which the free surface of rotating liquid helium II is studied. It is shown that the results, which are the same as for any ordinary liquid, are not in accordance with the hypothesis that the superfluid component remains stationary. On this basis an estimate of the critical velocity for superfluidity is made and is compared with previous determinations.

THE PROCEEDINGS OF THE PHYSICAL SOCIETY

Section B

VOL. 63, PART 9

1 September 1950

No. 369 B

Voltage Gradients in Long Gaseous Spark Channels

BY J. B. HIGHAM AND J. M. MEEK

Department of Electrical Engineering, University of Liverpool

MS. received 9th March 1950

ABSTRACT. Oscillographic measurements have been made of the variation with time of the voltage drops in long spark channels conducting aperiodic impulse currents rising to a peak in about $1/4 \mu\text{sec.}$ and decaying to half-value in from 10 to $85 \mu\text{sec.}$, with peak currents ranging from 60 to 700 amp. The lengths of the spark channels have been measured photographically, and the mean voltage gradients along the channels have been estimated. Results are given for sparks in air, nitrogen, oxygen and hydrogen at atmospheric and reduced pressures. Some preliminary measurements are described for sparks in water and oil, and for sparks in glass tubes. The voltage gradients in the various gases, with the exception of hydrogen, are independent of the peak current, but are influenced by the rate of current decay. The results are correlated with the current densities in spark channels, as recorded in a separate investigation.

§ 1. INTRODUCTION

THE growth of a spark is initiated by a streamer process, sometimes termed a leader stroke, which develops across the gap to form a conducting channel bridging the two electrodes. The leader stroke is followed by a main stroke, sometimes termed a return stroke, which completes the breakdown process, and the external circuit then discharges through the pre-ionized channel (Allibone and Meek 1938).

Before the development of the leader stroke the voltage across the spark gap has a high value, equal to or greater than the minimum breakdown voltage for the gap. On the occurrence of the main stroke the voltage falls to a low value which corresponds to the voltage drop along the spark channel itself. The purpose of the present investigation has been to measure this voltage drop and to determine its variation with time for different gaps in several gases under various circuit conditions. Some preliminary results have been referred to in an earlier publication (Meek 1947).

Previous experiments for gaps up to 1 cm. in length have been made by Craggs and Meek (1946), who showed that the voltage drop along sparks carrying 120 amp. for $3.6 \mu\text{sec.}$ in argon and in hydrogen at atmospheric pressure averages roughly 90 volts per cm.

In two oscillograms published by Flowers (1943), the voltage gradient for spark channels in air across a gap of 8.8 cm. is seen to be about 500 volt/cm. at a time of $1 \mu\text{sec.}$ after breakdown; and before the current has reached its peak amplitude. Also, in measurements by McCann and Clark (1943) of the

dielectric recovery characteristics of long air gaps, voltage gradients of 200 to 300 volts/cm. were recorded for sparks occurring along pre-ionized channels formed by earlier sparks. Norinder and Karsten (1949) record voltage gradients of the order of 100 volt/cm. in spark channels conducting oscillatory currents.

A complication in the earlier measurements with gaps of 1 cm. or less is that the voltage drops at the electrodes are comparable with the voltage drop in the discharge channel. A further complication is the diffusion of metal vapour into the gap, at a speed of about 10^5 cm/sec., so that the whole of the spark channel between electrodes spaced at 1 cm. contains both gas and metal vapour at times greater than about $5 \mu\text{sec.}$ from its initiation (Williams, Craggs and Hopwood 1949, Llewellyn-Jones 1945, 1946) and the voltage drop may therefore differ from that for a spark channel in the gas alone. Both these effects are of decreasing importance as the gap length is increased.

In the present investigation an attempt has been made to avoid the complications due to the electrodes by making measurements for longer gaps of length from 10 cm. up to 40 cm. Various gases have been studied, including air, nitrogen, oxygen and hydrogen at atmospheric and reduced pressures, and some preliminary measurements have been made for sparks in liquids with a shorter gap length.

§ 2. MEASUREMENT TECHNIQUE

Single impulse voltages were applied to the test spark gap from an 8-stage 1,000 kv. impulse generator of discharge capacitance $0.025 \mu\text{F.}$ The spark channel currents used had peak magnitudes between 60 and 700 amp., with various rates of decay determined by the series resistor R_s , as shown in Figure 1. The

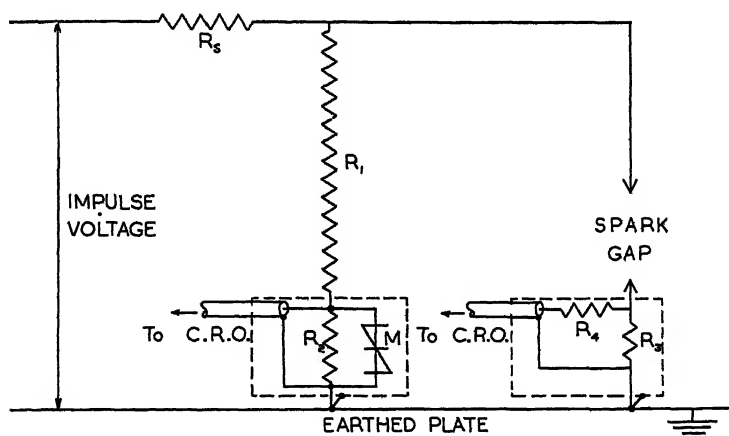


Figure 1. Circuit for voltage and current measurements.

currents reached their peak values in approximately $0.25 \mu\text{sec.}$, and the values of R_s were chosen so that in general the currents decayed to half value in either 10 or $28 \mu\text{sec.}$ Voltages across the channels were measured at times ranging between 0.25 and $80 \mu\text{sec.}$ after breakdown. The times interval between successive sparks was so long that each spark was uninfluenced by the preceding one.

Oscillographic measurements of voltage and current were made with a 10 kv. sealed-off cathode-ray oscillograph (Prime and Saxe 1949) the screen of which was

photographed on 35 mm. film with a 2-inch $f/1.0$ bloomed lens. Time sweep durations down to $0.5 \mu\text{sec.}$ were available.

In order that the voltage across the spark gap could be recorded on the oscillograph, a potential divider had to be used, the dividing ratio of which was governed by the consideration that the voltage drop across the spark channel had to produce a measurable deflection on the oscillograph. If this is done with a normal resistance divider, the voltage across the gap before the gap breaks down is so high that it can cause damage to the oscillograph. For this reason a potential divider was constructed as shown schematically in Figure 1, with a non-linear resistor shunting the lower resistor R_2 of the divider.

Both high and low voltage resistance elements R_1 and R_2 were made from resistance ribbon (in which the warp is of silk and the weft of constantan wire) which has a reasonably low uniform inductance of about $1 \mu\text{H.}$ per inch. R_2 had a value of about 65 ohms, matching the surge impedance of the 12 ft. long polythene insulated concentric cable enclosed in conduit leading to the oscillograph. Shunting R_2 was a disc of Metrosil non-linear resistance material, the characteristic of which may be expressed in the form $V = KI^\beta$ where V = voltage, I = current, K = a constant, and the index β is about 0.25. The sample used was so chosen that its effective resistance is large compared with R_2 when the comparatively low post-breakdown voltage across a spark channel is being recorded; thus the divider is linear in its response to such voltages, but when the pre-breakdown voltage is applied, the effective resistance of the Metrosil disc is low enough to limit the divider output voltage to a safe value. Metrosil responds to transients without appreciable time lag, a consideration which precludes the use of a glow-tube or a similar safety device. Tests were made with and without the semiconductor, which showed that its presence did not effect the response of the potential divider to voltage transients similar to those being measured. A range of low value 'non-inductive' resistance units, R_3 , were available for insertion in the circuit when measurements of current were required. A typical oscillogram of the residual voltage drop across a spark channel air is shown in Figure 2 (Plate I).

The usual precautions adopted in high-voltage measurements were adhered to throughout the work, and at least eight oscillograms were taken for any one condition being investigated. The oscillograms were then enlarged by about eight diameters, and tracings were made which were analysed with a specially photographed graticule. It was found that the axes of the cathode-ray oscillograph were 3° from normal, and the graticule was constructed accordingly. The zero for measurements of time from breakdown was taken where the rapidly decaying initial portion of the voltage oscillogram is asymptotic to the Y axis of the graticule, as illustrated in Figure 2. The possible error involved is not more than about $0.1 \mu\text{sec.}$

§ 3. SPARK LENGTHS

The voltage drops measured across successive spark channels were found to have a 'scatter' not accountable for by measurement errors. It was realized early in the investigation that this was probably caused by variations in channel lengths, and consequently an attempt was made to record these lengths.

The path of any particular spark channel can be determined by taking photographs from two directions at right angles, but the subsequent analysis involved is prohibitive for a large number of records. A fairly close estimate of the total

length may be obtained on a statistical basis from photographs taken with one camera, and this procedure has been adopted in the present investigation.

The estimated mean total spark length, L , may be expressed as

$$L = L_g + k(L_p - L_g),$$

where L_g is the gap length, L_p the mean photographed or projected length, and k a factor to be determined.

From a consideration of hypothetical spark formations, such as semicircular in a plane and helical in three dimensions, and allowing for small irregularities, the authors have assigned the value 2.0 to the factor k . However, the value of k is not critical; for example, if the mean photographed length is 10% greater than the gap length and the correct value of k is 2.0, the error involved by taking values of $k = 2.3$ and 1.7 is only $\pm 2\frac{1}{2}\%$ of the correct length.

Photographs were taken of up to 50 sparks under each set of conditions of gap length, circuit constants, current, gas and gas pressure employed in the voltage measurements. Typical photographs are shown in Figure 3 (Plate II), where the changes in spark length for the different conditions are readily apparent. The spark channels in hydrogen are particularly irregular, and, for an applied voltage of approximately twice the minimum breakdown voltage, the mean total spark length is 68% greater than the gap length. The corresponding value measured for air is only 10%. For other conditions investigated, the lengths measured in hydrogen were more than in the other gases but not by so great an amount.

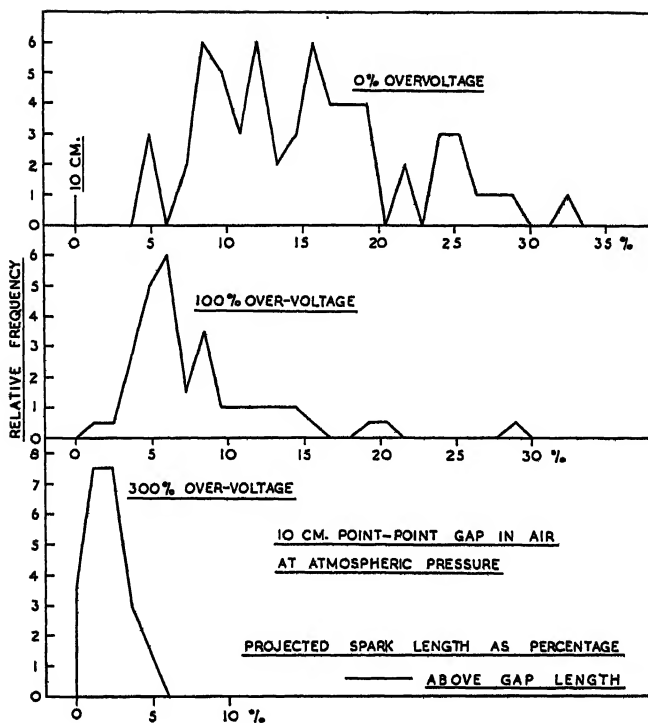
In Figures 4(a) and 4(b), the photographed spark lengths, expressed as a percentage above the gap lengths, are plotted against their frequency of occurrence for particular conditions. The resultant distribution curves illustrate the general tendency for the spark lengths to diminish and to become more uniform as the applied voltage is increased, or as the gas pressure is reduced with the applied voltage kept constant.

This is clearly shown in the lower graph of Figure. 5, in which the mean total spark length is plotted as a function of gap length for the case where the applied voltage was directly proportional to the gap length. The two curves correspond to the first two columns of Table 1 (see §4.1). In curve A, the voltage applied to the 10 cm. gap was little more than that required for breakdown, and the resulting mean spark length is high. As the minimum breakdown voltage for long point-point gaps does not increase linearly with gap length, the gaps longer than 10 cm. were subjected to voltages appreciably higher than the minimum breakdown voltages. Thus the curves bear out the general finding that the spark length diminishes with increasing over-voltage.

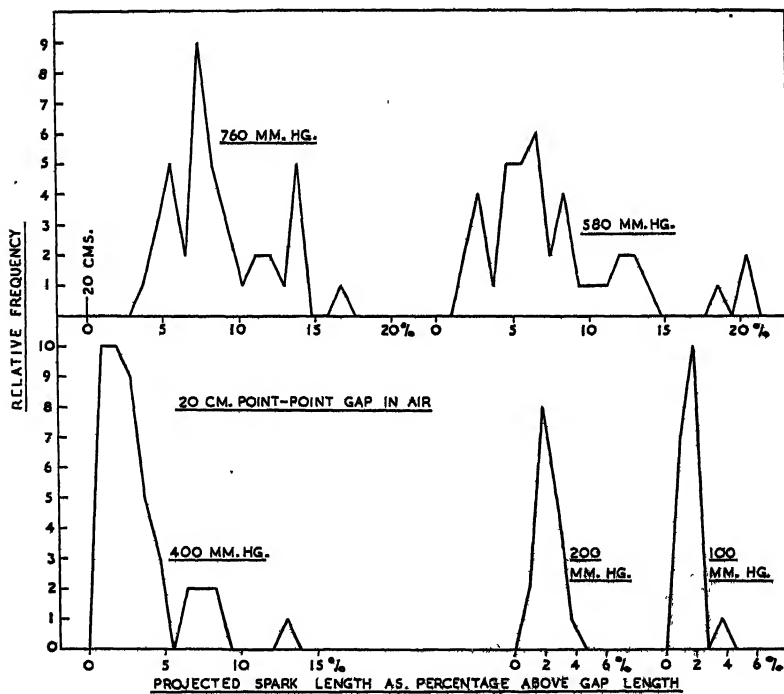
§4. VOLTAGE GRADIENTS

For any one set of conditions being investigated, several oscillograms have been obtained to show the total voltage drop across the spark gap as a function of time, and from the analysis of these oscillograms a mean curve has been drawn showing the variation with time of the ratio of the mean voltage drop to the mean measured length L of the spark. This curve may then be considered to show the mean voltage gradient along the spark channel as a function of time.

This method of determining the voltage gradient rests on the assumption that the gradient is uniform along the spark channel, neglecting electrode effects, and



(a)



(b)

e 4. Graphs showing the relative frequency of occurrence of spark lengths, expressed as a percentage above the gap length :

- (a) For various applied voltages at constant gas pressure.
(b) For various gas pressures at constant applied voltage.

seems to be justified by the results obtained. The graph of Figure 6, in which the mean voltage across the spark channel is plotted as function of gap length at various times from breakdown for a particular current characteristic, illustrates this linear relation and indicates that the voltage drop at the electrodes and in the electrode connections is small compared with the voltage drop across the spark.

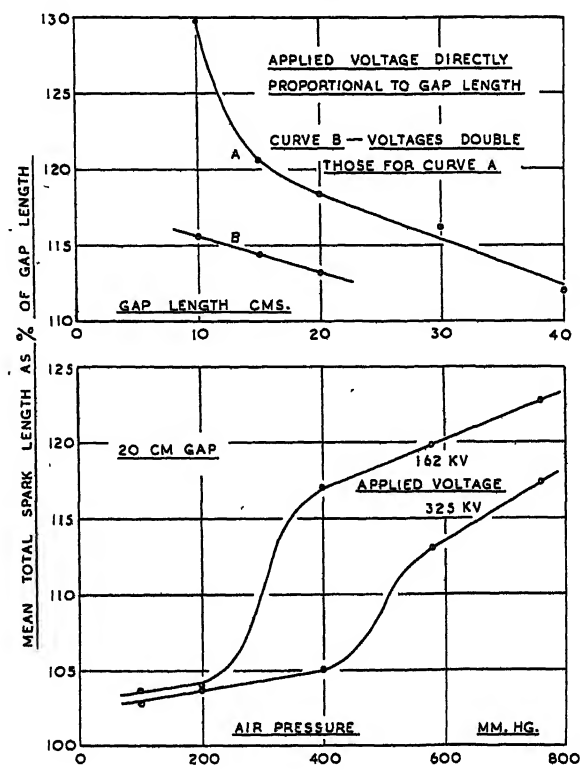


Figure 5. The dependence of spark length on applied voltage and gas pressure.

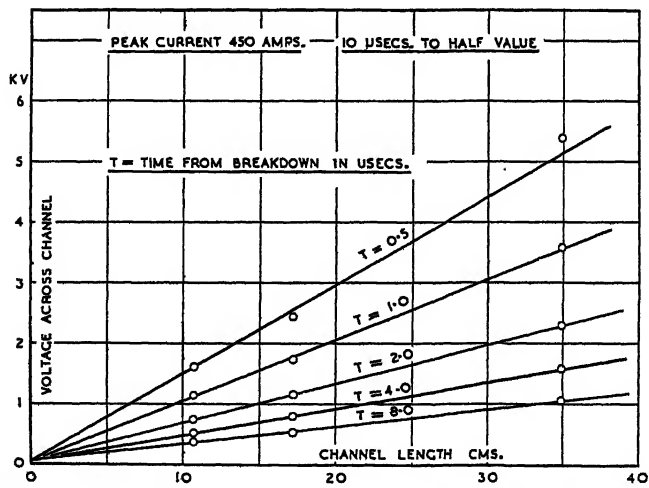


Figure 6. The linear relation between voltage drop and channel length at various times from breakdown.

As a further check on the effect of electrode connections, voltage oscillograms were recorded with the usual potential divider with the spark gap short circuited. Apart from small initial oscillations dying out in less than $0.5 \mu\text{sec.}$ the result was a straight trace coincident with the zero line.

4.1. Air at Atmospheric Pressure

Some of the results obtained are shown in Tables 1 and 2 and apply to spark channels between point electrodes, one of which is earthed, when the high voltage electrode is subjected to impulse voltages of positive polarity through series resistors of 750 and 2,000 ohms respectively. The currents are the values corresponding to the related sets of voltage measurements at various times from breakdown. These are taken from current oscillograms such as those reproduced

Table 1

(1) Gap (cm.)	(2) $\mu\text{sec.}^*$	(3) kv.	(4) v/cm.	(5) amp.	(6) kv.	(7) v/cm.	(8) amp.	(9) kv.	(10) v/cm.	(11) amp.
10	0.5	2.15	165	150	1.63	142	300	1.62	152	450
	1	1.41	108	145	1.15	100	290	1.15	107	436
	2	0.95	73	135	0.84	73	270	0.75	70	406
	4	0.60	46	118	0.60	52	236	0.53	49.5	350
	8	0.45	34.5	89	0.42	36.5	178	0.39	36.5	266
	16	0.32	24.5	50	0.28	24.5	100			
	24	0.24	18	29	0.20	17.5	58			
15	0.5	3.11	173	225	2.43	142	450	2.70	164	675
	1	2.01	112	218	1.72	100	436	1.80	109	654
	2	1.30	72	203	1.15	67	406	1.14	69	609
	4	0.85	47	175	0.79	46	350	0.82	50	525
	8	0.60	33	133	0.54	31.5	266			
	16	0.45	25	75	0.38	22	150			
	24	0.32	17.5	44	0.26	15	88			
20	0.5	3.54	150	300	3.14	139	600			
	1	2.38	101	290	2.18	96	580			
	2	1.60	68	270	1.50	66	540			
	4	1.14	48	236	1.06	47	472			
	8	0.78	33	178	0.76	33.5	356			
	16	0.56	24	100	0.53	23.5	200			
	24	0.38	16	58	0.34	15	116			
30	0.5	5.40	155	450						
	1	3.60	103	436						
	2	2.31	66	406						
	4	1.60	47	350						
	8	1.08	31	266						
	16	0.78	22.5	150						
	24	0.52	15	88						
40	0.5	6.48	145	600						
	1	4.64	104	580						
	2	3.00	67	540						
	4	2.04	45.5	473						
	8	1.32	29.5	356						

0.025 $\mu\text{F.}$ impulse generator.
Series resistance 750 ohms.
Air at atmospheric pressure.
Current decaying to half value in 10 $\mu\text{sec.}$

* Column (2) gives time from breakdown.

in Figure 7 (Plate I). Columns (3), (6) and (9) give the mean total voltage measured across the sparks at various times from breakdown. The adjacent columns (4), (7) and (10) give the mean voltage gradient along the spark channel obtained by dividing the kilovolt values by the corresponding mean spark lengths.

For either current waveform there is close agreement in the magnitude of the voltage gradients for the various gap lengths and peak currents. For example, referring to Table 1, for gap lengths of 10 to 40 cm. and peak currents of 150 to

675 amp. the voltage gradients obtained 2 μ sec. after breakdown lie in the range of 66 to 73 volt/cm. This is almost certainly within the limits of overall experimental accuracy.

A comparison of Tables 1 and 2 indicates the general effect of current duration. This effect is shown more clearly in Figure 8 in which voltage gradients for various

Table 2

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Gap (cm.)	μ sec.*	kv.	v/cm.	amp.	kv.	v/cm.	amp.	kv.	v/cm.	amp.
10	0.5	1.80	145	60	1.45	130	125	1.62	151	190
	1	1.37	110	59	1.13	101	124	1.17	109	188
	2	1.04	84	58	0.87	78	121	0.83	78	183
	4	0.78	63	55	0.62	55.5	115	0.61	57	175
	8	0.54	43.5	50	0.48	43	105	0.49	46	160
	16				0.37	33	86			
	24				0.27	24	70			
20	0.5	3.44	150	125	2.74	125	250			
	1	2.46	107	124	1.98	90	248			
	2	1.74	76	121	1.41	64.5	242			
	4	1.26	55	115	1.05	48	230			
	8	0.96	42	105	0.85	39	210			
	16	0.74	32	86	0.68	31	172			
	24	0.60	26	70	0.55	25	140			
40	0.5	6.00	136	250						
	1	4.20	95	248						
	2	2.96	67	242						
	4	2.11	48	230						
	8	1.63	37	210						
	16	1.32	30	172						
	24	1.10	25	140						

0.025 μ F. impulse generator.
Series resistance 2,000 ohms.
Air at atmosphere pressure.
Current decaying to half value in 28 μ sec.

* Column (2) gives time from breakdown.

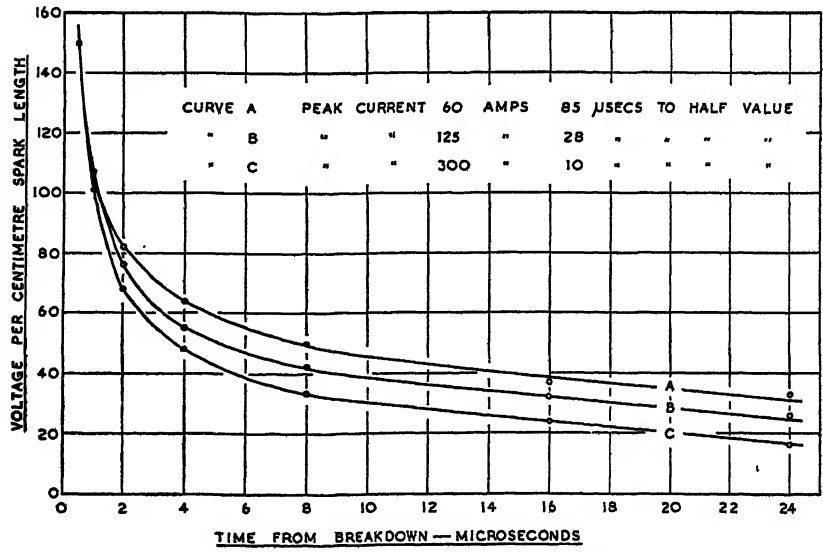


Figure 8. Voltage gradients in air for three different rates of current decay.

current durations are plotted. The lower the rate of current decay, the higher the voltage maintained across the spark channel.

A few measurements were made for point-point and point-plane gaps with positive and negative polarity voltages applied. The results were practically

identical and showed that the gap geometry does not affect the magnitude of the voltage drop. No differences were observed in the results obtained for electrodes of different materials, such as steel and brass, though such differences may be expected for measurements with short gaps when metal vapour is present in all parts of the discharge and not just in the region of the electrodes.

4.2. Voltage Gradients in Nitrogen, Oxygen and Hydrogen at Atmospheric Pressure

All the measurements were made for sparks across a 20 cm. gap between pointed electrodes, in a Perspex cylinder about 35 cm. in diameter and 90 cm. long. The gases used were of standard commercial purity.

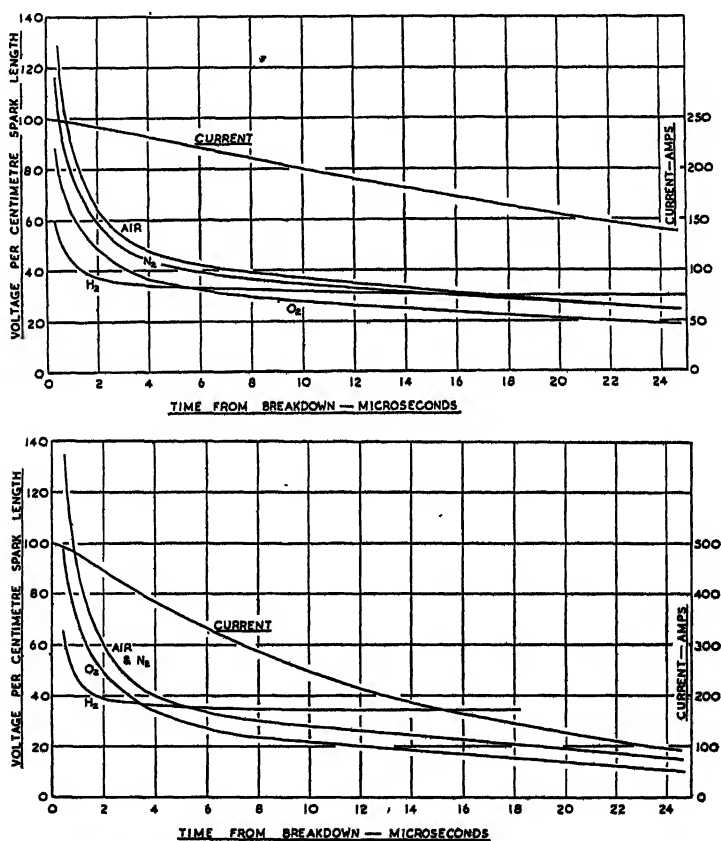


Figure 9. Voltage gradients in air, N_2 , O_2 , and H_2 with the associated current waveforms.

Figure 9 shows typical curves of voltage gradient as a function of time for nitrogen, oxygen, hydrogen and air. The characteristics for air, nitrogen and oxygen were found to be very similar in form, air always having the highest voltage gradient and oxygen the lowest, for peak values of current ranging between 125 and 500 amp. and for durations of 10 and 28 μ sec. to half value. On the other hand, hydrogen always had an initially lower voltage gradient than those of the other gases but after two or three microseconds the voltage ceased to fall so rapidly and quickly reached a minimum. This is illustrated in the oscillograms of Figure 10.

(Plate I) which show the recovery of voltage across the gap and the corresponding decay of the current to zero. Sparks in hydrogen also differed from those in the other gases in that the voltage gradient showed a dependence on the peak value of the current. For example, with currents decaying to half value in $10\ \mu\text{sec.}$ the voltage gradients from $0.25\ \mu\text{sec.}$ after breakdown to the time of commencement of voltage recovery were 40% greater for a peak current of 250 amp. than for a peak current of 500 amp.

4.3. Voltages Gradients in Gases at Reduced Pressure

A comprehensive series of measurements was made on sparks in air, nitrogen and oxygen at pressures from atmospheric down to 100 mm. Hg. Typical photographs of sparks in this pressure range are shown in Figure 3 which illustrates the general tendency for the spark channel to become straighter with reduced pressure (see also the graphs of Figure 4 (b)). At 100 mm. Hg and particularly at lower pressures still, the spark is diffuse and the light emitted is relatively weak.

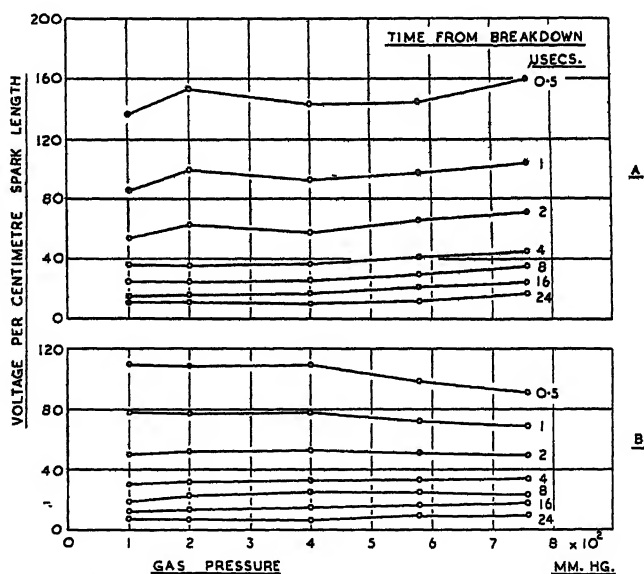


Figure 12. The effect of gas pressure on the voltage gradient. Current decays to half value in $10\ \mu\text{sec.}$

A. 20 cm. gap in air.

B. 20 cm. gap in oxygen.

Also the instant of breakdown is ill-defined due to the comparatively high current flowing during the formative stages of the spark (see oscillogram of Figure 11, Plate I) which produces a voltage drop in the series resistor. For this reason, 100 mm. Hg was the lower limit of pressure for which useful voltage measurements were made.

The curves of Figure 12, in which voltage gradients at various times from breakdown are plotted against gas pressure, are typical of the results obtained, and show that variations of gas pressure, within the range investigated, has little effect on the voltage gradient.

4.4. Sparks of Restricted Cross Section

A series of voltage measurements was made on sparks in glass tubes and between glass plates.

Thick walled tubes, of Pyrex glass, with bores ranging between 1 and 6 mm. were used in the experiments. Sparks about 15 cm. in length were produced between steel wire electrodes, inserted loosely one at each end of the tube in use, and the voltage across the spark channel was measured. Typical voltage oscillograms are shown in Figure 13 (Plate I) and are characterized by an initial voltage decay, similar to that for an unrestricted spark, followed by a minimum and a subsequent rise in voltage. The initial region of voltage decay is not greatly affected by the presence of the glass tube, as indicated by the results of Figure 14.

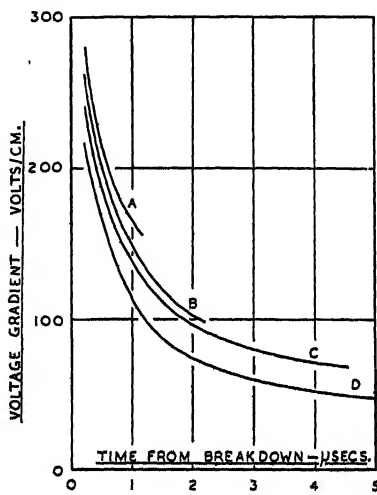


Figure 14. Initial voltages across spark channels restricted in glass tubes with the corresponding result in free air for comparison.

- A. Tube bore 0.197 cm.
- B. Tube bore 0.385 cm.
- C. Tube bore 0.57 cm.
- D. Free air.

It is reasonable to assume that the minimum in the voltage curve is related to the time when the spark channel has expanded to fill the tube, so that measurements with various tube bores should give an indication of the rate of expansion of the spark channel. The results for two currents are shown in Figure 15, curves A and B, the minima for the higher current being reached more quickly than for the lower current. The time was measured to the point where the voltage has decayed to within 5% of the minimum.

Voltage drops were also measured in sparks restricted between glass plates. It was found that the minimum occurred, but after a time greater than that required in the tubes. A result is shown in Figure 15, curve C. It is possible that the greater time arises because lateral expansion is possible in one direction after expansion has ceased in the other. As for glass tubes, a higher current produces a minimum in the voltage curve after a shorter time from breakdown.

The curves D and E of Figure 15 refer to spark channel diameter measurements on unrestricted spark channels made with the rotating mirror camera (Higham and Meek 1950) for the same two peak currents used in the glass tube experiments.

While the initial rate of expansion measured by the two methods is of the same order, the rate determined by the glass tube technique in the later stages is very much greater. According to the photographic results a spark channel conducting a peak current of 450 amp. reaches a maximum diameter of little more than 4 mm., whereas the tube measurements indicate that this diameter is reached in about $1.5 \mu\text{sec.}$ and that the expansion continues after this time at a nearly uniform rate. One possible explanation of the observed discrepancy between the results obtained by the two techniques is that the boundary of the visual photographed section of the channel and the boundary of the ionized region may not coincide. Curves A

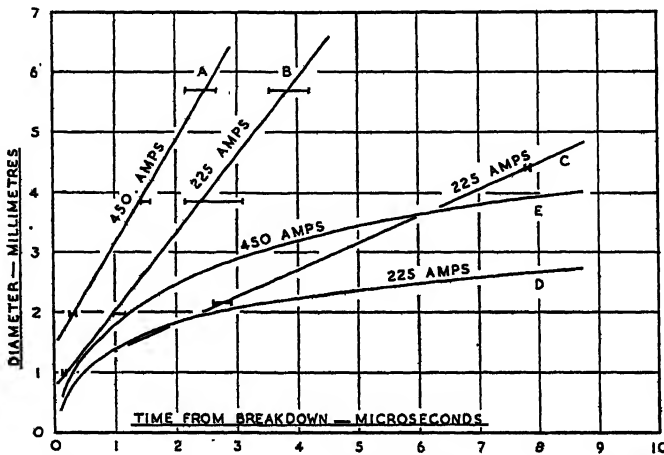


Figure 15. The expansion of spark channels in air as interpreted from voltage measurements on restricted channels. The results obtained photographically are included for comparison.

and B relate to the condition where the electrical characteristics of the spark become affected by the presence of the tube, and hence may be assumed to give the boundary of the ionized region, whereas curves A and E give the photographed boundary. According to this interpretation the effective diameter of the ionized region of a spark channel carrying 450 amp. at $2 \mu\text{sec.}$ after breakdown is 4.9 mm. while the effective diameter of the visible channel is only 2.4 mm. However, the presence of the tube may itself influence the expansion rate of the channel and further experimental study is required to confirm, or reject, the above proposals.

4.5. Voltage Gradients in Spark Channels in Liquids

A few measurements were made to determine the order of the voltage gradient for spark channels in liquids. Table 3 summarizes some results obtained for spark channels in ordinary tap water between brass pointed electrodes with a gap

Table 3. Voltage Gradients in volt/cm.

Peak current (amp.)	Time from breakdown ($\mu\text{sec.}$)						
	0.5	1	2	4	8	16	24
250	1700	1110	700	490	300	230	130
300	1280	920	630	420	300	240	160
750	1270	830	570	390	300	230	120

length of 2.5 cm. and currents decaying to half value in 10 μ sec. No appreciable difference was recorded between the results for distilled and for ordinary tap water. The spark channel voltage gradients in transformer oil are of the same order as those in water.

The voltage gradients given in Table 3 denote the measured voltage drops divided by the length of the spark gap. It is probable that the sparks have a greater length than the gap, so that the true voltage gradients may be slightly lower than those given in the Table. If, as is the case for gaseous spark channels, the lengths of the sparks decrease with increasing over-voltage, the true voltage gradient for the currents investigated would be about 800 volt/cm. at 1 μ sec. after breakdown and about 290 volt/cm. at 8 μ sec. after breakdown. These values are about eight times those for air at atmospheric pressure.

§ 5. DISCUSSION

The present research concerning the voltage gradient along spark channels is one of a series of investigations into the spark channel now being performed in these laboratories. These investigations include measurements of the variation with time of the effective diameter of the spark channel and the radial distribution of light emission, and also spectroscopic studies from which ion densities may be determined. Until all the results of the various interrelated researches are known the interpretation of any single study is handicapped, particularly as there is little previous information available concerning the discharge processes involved. A complete theoretical treatment is therefore not attempted in this discussion which is intended primarily to draw attention to certain features of the results and to show some general correlation between them.

The change recorded in the length of the spark channel with variations in the magnitude of the applied voltage, for a fixed gap between the electrodes, is an effect which does not appear to have been noted previously. As the path taken by the leader stroke determines the shape of the spark channel (Allibone and Meek 1938), the changes in channel length can be explained only by consideration of the leader stroke process. No complete theoretical analysis of the mechanism of the leader stroke is yet available (see, however, Schonland 1938, Meek 1939, Bruce 1944, Loeb 1948), but in general it is assumed that the leader stroke proceeds as a consequence of electron avalanche development in the region of its tip. As the electron avalanches are initiated in a random manner, and can grow within a large solid angle surrounding the tip of the leader stroke, the leader stroke develops along a zig-zag path.

If the voltage applied to the gap is increased above the minimum value required to cause breakdown, the number and size of electron avalanches produced in the region surrounding the tip may be expected to increase and there is an enhanced probability of successful avalanche development along the line of greatest field strength between the tip of the leader stroke and the opposite electrode.

With reduction in gas pressure, the mean free path of electrons is increased, so that the electron avalanches are longer, and as the leader stroke is made up of a succession of avalanches it may itself be expected to take a less irregular course. Also, at lower gas pressures there is greater diffusion during the growth of avalanches and streamers or leader strokes, and the path of the leader stroke assumes a smoother and more rounded form. The more direct path taken at reduced pressures would also seem to imply an increase in the number of

avalanches produced round the tip of the leader stroke, so that there is a greater chance of successful avalanche growth along the line between the tip and the opposite electrode.

An increase in spark length at pressures higher than atmospheric has been recorded by Howell (1939) for breakdown between a point and a plane. In his experiments an A.C. voltage was used and considerable space charges were produced before breakdown occurred, with consequent field distortion which may have affected the leader stroke development. In the present experiments, the time between the application of the impulse voltage and the breakdown of the gap is too short for drift of space charge across the gap.

All the present measurements have been made for point-point gaps, for which simultaneous growth occurs of leader strokes from both positive and negative points (Allibone and Meek 1938). Differences may be expected for point-point gaps, with points of positive or negative polarity, or for sphere-sphere gaps, and further relevant studies are being undertaken.

The voltage drop varies directly with the gap length and hence the voltage gradient along the spark channel may be assumed to be nearly uniform. There is no reason to suppose that one portion of the spark channel should differ from another, except in the region of the electrodes, or where the direction of the spark path is sharply inclined to the line joining the two electrodes. For instance, the gradient at bends of the channel may differ from that of adjacent portions of the channel, but on balance these variations probably cancel out and give a resultant average voltage gradient for the whole channel varying little from spark to spark.

The magnitudes of the voltage gradients recorded in the various gases are appreciably higher than those for arcs carrying the same current. Measurements by von Engel (1929) for an arc in air, 10 cm. in length and carrying a current of several hundred amperes, show that the voltage gradient is about 2 volt/cm. The present results for sparks carrying a current of approximately the same value give a voltage gradient of about 100 volts/cm. at a time of $1\text{ }\mu\text{sec.}$ after breakdown and about 40 volt/cm. at a time of $8\text{ }\mu\text{sec.}$

It is clear that a higher voltage gradient would be expected in a spark than in an arc, as in the former case equilibrium has not been established. In a spark channel the gas density is initially the same as that of the surrounding gas, but with the sudden input of energy the effective gas temperature rises rapidly, to a value thought to be of the order of $10,000^{\circ}\text{C.}$ (Craggs and Meek 1946), with a consequent increase in gas pressure and a rapid expansion of the channel. This expansion continues, and conditions within the channel change, until the gas pressure within the channel falls to that of the surrounding gas. Meanwhile, if the current is maintained, the gas density falls and the temperature assumes a value of about $6,000^{\circ}\text{C.}$ for arcs in air at atmospheric pressure. On the above basis it is possible, in some respects, to liken the spark to the case of an arc which is artificially constricted to a cross section appreciably smaller than the normal value.

In the light of the above general picture of the spark it is interesting to draw attention to some of the trends observed in the present measurements. One distinctive feature is that over a wide range of currents there is little change in the value of the voltage gradient at given times from breakdown. For instance, as shown in Table 1, for a gap of 10 cm. in air, the voltage gradient in channels carrying a peak current ranging from 150 to 450 amp., and decaying to half value in $10\text{ }\mu\text{sec.}$

varies between 142 and 165 volt/cm. at 0.5 μ sec. after breakdown, and between 34.5 and 36.5 volt/cm. at 8 μ sec. after breakdown. Measurements of spark channel diameters (Higham and Meek 1950) indicate that at 8 μ sec. after breakdown the corresponding sparks, carrying peak currents of 150 and 450 amp., have diameters of 0.21 and 0.39 cm. respectively. The average current densities in these two cases are then 2.5 and 2.2 ka/cm² and are hence roughly equal. From analysis of these and other records of spark channels in air, it appears to be true, to a first approximation, that at a given time after breakdown the voltage gradients for sparks carrying different currents are the same and, further, that the current densities at a given time are the same. Consequently, for a given current wave shape, the conductivity of a spark channel is roughly constant at a given time from breakdown for a wide range of peak currents, between 100 and 1,000 amp. approximately. For currents rising to peak value in about 0.25 μ sec., and decaying to half value in times longer than 10 μ sec. the value obtained for the conductivity is about 150 ohm⁻¹ cm⁻¹ at a time of 1 μ sec. from breakdown. At a time of 8 μ sec. from breakdown the value is more dependent on current wave shape and values of 75 ohm⁻¹ cm⁻¹ and 90 ohm⁻¹ cm⁻¹ were obtained for currents decaying to half value in 10 μ sec. and 28 μ sec. respectively. Norinder and Karsten (1949) quote a conductivity of 400 ohm⁻¹ cm⁻¹ for sparks carrying oscillatory currents of peak amplitude between 10 and 100 kiloamp.

The product of current density and voltage gradient gives the power input per unit volume of the spark channel. For the spark channels in air the power input at 1 μ sec. after breakdown is of the order of 1 megawatt per cm³.

The close connection between voltage gradient and channel diameter is also shown by the results for gases other than air. The voltage gradients recorded in nitrogen are closely the same as for air and the channel diameter expands at a similar rate (Higham and Meek 1950). In oxygen the voltage gradient is slightly lower than in air but the channel diameter is greater, and consequently the current density is lower. In hydrogen the voltage gradient is lower than in air for the early stages of the spark channel, and is higher in the later stages; corresponding observations of the rate of expansion of the channel (Higham and Meek 1950) show that the expansion is more rapid in the early stages and slower in the later stages.

The results obtained for the sparks in tubes show that the voltage gradient is increased by the artificial constriction of the channel cross section. Even before the spark channel has expanded to fill the tube, the walls may be expected to influence the characteristics of the spark channel appreciably, partly because of their effect in the rate of energy loss but also because of the reflections of the pressure wave generated when the spark is initiated. For such reasons it is unwise to rely too greatly on any deductions concerning the rate of growth of channel diameters from a comparison between the voltage drops for sparks in free air and for sparks in tubes. The comparatively high gradients in spark channels in liquids are probably caused largely by the constriction of the channel, which cannot expand freely because of the opposition from the liquid boundary.

The voltage gradient measurements of Figure 8, which show the effect of changing the rate of current decay, have been extrapolated to determine the voltage gradients for a unit function current, i.e. a current rising to its peak magnitude instantaneously and thereafter remaining constant. The results are given in

Table 4, and may be expected to apply with reasonable accuracy to spark channels in air at atmospheric pressure when conducting unit function currents of magnitudes ranging between 50 to 1,000 amp.

Table 4

Time from breakdown ($\mu\text{sec.}$)	0.5	1	2	4	8	16	24
Voltage gradient (volt/cm.)	150	118	93	72	56	43	35

In future work it is intended to overcome the complication of current decay by the use of square current pulses which rise to a peak in a time of the order of $0.1 \mu\text{sec.}$ and remain at that value for a prescribed period, of possibly $10 \mu\text{sec.}$ or more. Experiments with square pulses have been made by various investigators, but in all cases the voltage available was such that it was not possible to obtain breakdown of gaps of more than about 2 cm. With the impulse generator used in the present experiments, the voltage available, of 1 million volts, is sufficient to produce breakdown gaps up to about 150 cm. but it is not possible to produce a square current pulse of amplitude 1,000 amp. or more in the spark channel. A new impulse generator is now being built with which it is hoped to produce pulses of 100 kv. giving a current of, for example, 1,000 amp. lasting for $10 \mu\text{sec.}$

A further improvement in technique may also be adopted in future investigations in order to determine more accurately the voltage drop in the early stages of the channel, within $0.2 \mu\text{sec.}$ from breakdown. The actual instant of breakdown is not readily recorded to within less than $0.1 \mu\text{sec.}$ because of the character of the oscillogram, and consequently the measurement of the voltage drops in the early stages of spark formation is inaccurate. One possible method of improving the records is to use two synchronized oscillographs, one within a linear divider and the other with a non-linear divider. From the former a more precise record of the breakdown time can be obtained, while from the latter the voltage drop may be measured.

REFERENCES

- ALLIBONE, T. E., and MEEK, J. M., 1938, *Proc. Roy. Soc. A*, **166**, 97, 169, 246.
 BRUCE, C. E. R., 1944, *Proc. Roy. Soc. A*, **183**, 228.
 CRAGGS, J. D., and MEEK, J. M., 1946, *Proc. Roy. Soc. A*, **186**, 241.
 VON ENGEL, A., 1929, *Z. Tech. Phys.*, **10**, 505.
 FLOWERS, J. W., 1943, *Phys. Rev.*, **64**, 225.
 HIGHAM, J. B., and MEEK, J. M., 1950, *Proc. Phys. Soc. B*, **63**, 649.
 HOWELL, A. H., 1939, *Trans. Amer. Inst. Elect. Engrs.*, **58**, 193.
 JONES, F. LLEWELLYN, 1945, *J. Soc. Chem. Ind.*, **64**, 317; 1946, *Nature, Lond.*, **157**, 298.
 LOEB, L. B., 1948, *J. Franklin Inst.*, **246**, 123.
 McCANN, G. D., and CLARK, J. J., 1943, *Trans. Amer. Inst. Elect. Engrs.*, **62**, 45.
 MEEK, J. M., 1939, *Phys. Rev.*, **55**, 972; 1947, *Nature, Lond.*, **160**, 110.
 NORINDER, H., and KARSTEN, O., 1949, *Ark. Mat. Astr. Fys.*, **36A**, Part 4.
 PRIME, H. A., and SAXE, R. F., 1949, *Proc. Instn. Elect. Engrs.*, **96**, Pt. II, 662.
 SCHONLAND, B. F. J., 1938, *Proc. Roy. Soc. A*, **118**, 233.
 WILLIAMS, G. C., CRAGGS, J. D., and HOPWOOD, W., 1949, *Proc. Phys. Soc. B*, **62**, 49.

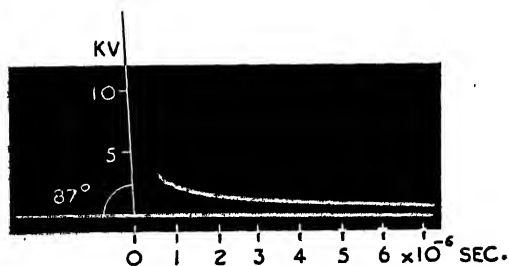


Figure 2. Oscilloscope of voltage drop across spark channel on which the Y axis has been drawn. (The original negative shows a clearly defined trace above 10 kv.)

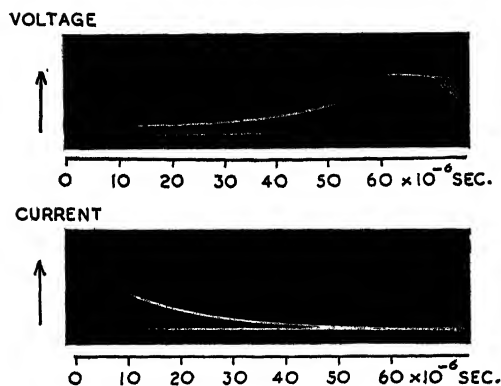


Figure 10. Oscilloscope showing the recovery of voltage across a hydrogen spark channel with the corresponding decay of current to zero. After current extinction the voltage decays by current flowing through the potential divider in parallel with the spark gap.

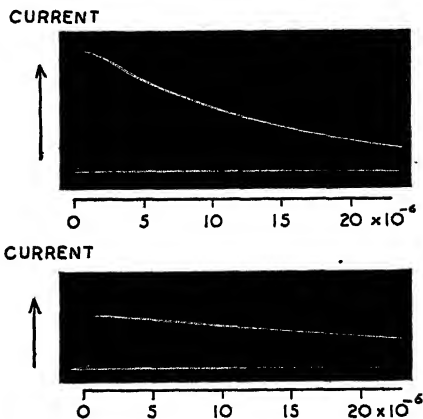


Figure 7. Typical current oscilloscope
A. Current decaying to half value in 10μ with 750 ohms series resistance.
B. Current decaying to half value in 28μ with 2,000 ohms series resistance.



Figure 11. Oscilloscope showing the current flowing through a spark channel in a at 200 mm. Hg. (An appreciable current flows before complete breakdown occurs at 1.8×10^{-6} sec.)

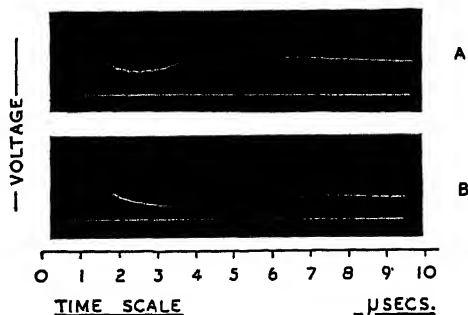


Figure 13. Oscilloscope showing the voltages across spark channels restricted in glass tubes. A. 0.197 bore. B. 0.385 cm. bore.

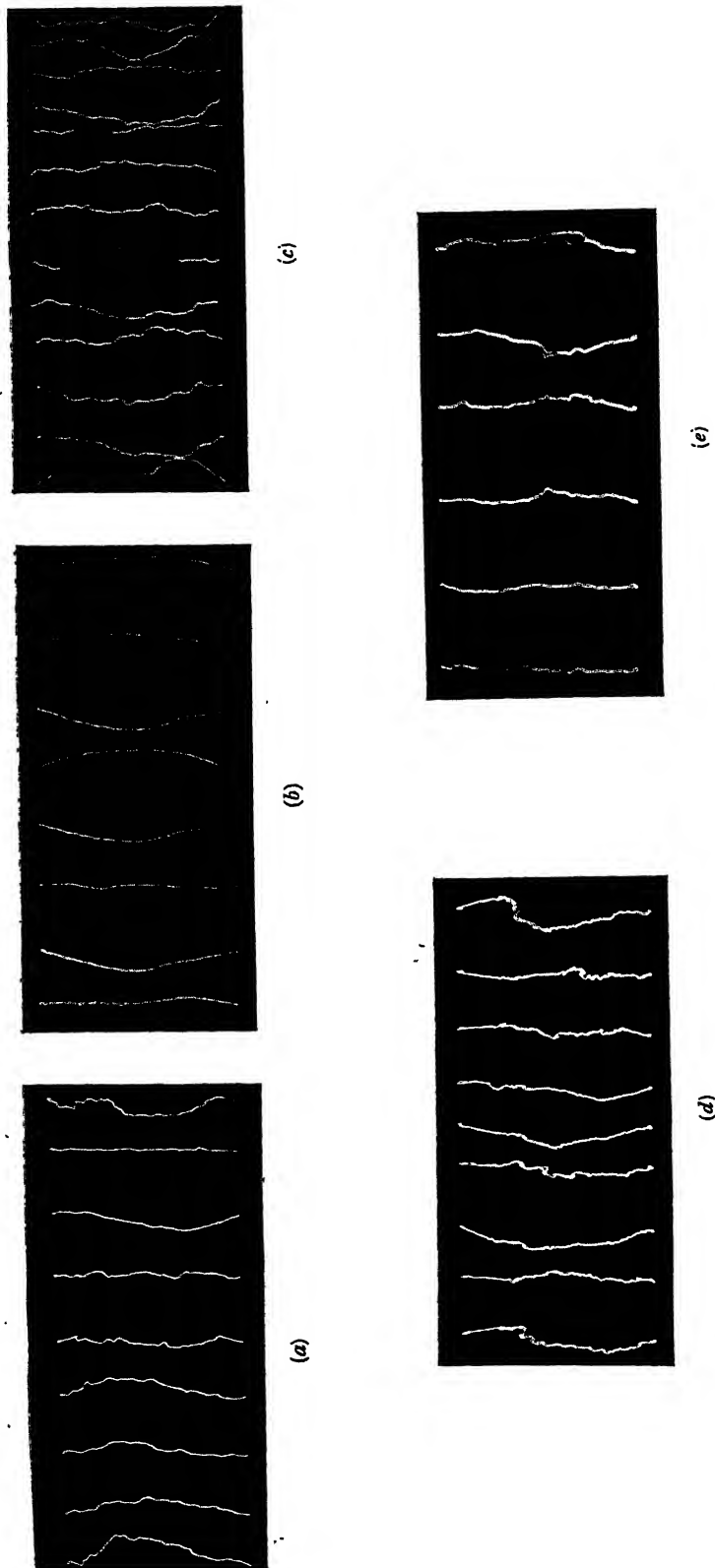


Figure 3. Typical photographs of sparks in 20 cm. gaps.

- (a) Nitrogen at atmos. pressure. Applied voltage 200 kv. Peak current 250 amp.
- (b) Nitrogen at 200 mm. Hg. Applied voltage 200 kv. Peak current 250 amp.
- (c) Air at atmos. pressure. Applied voltage 200 kv. Peak current 250 amp.
- (d) Hydrogen at atmos. pressure. Applied voltage 200 kv. Peak current 250 amp.
- (e) Oxygen at atmos. pressure. Applied voltage 490 kv. Peak current 500 amp.

(The second images in these photographs are caused by reflection from the wall of the cylindrical Perspex gas chamber.)

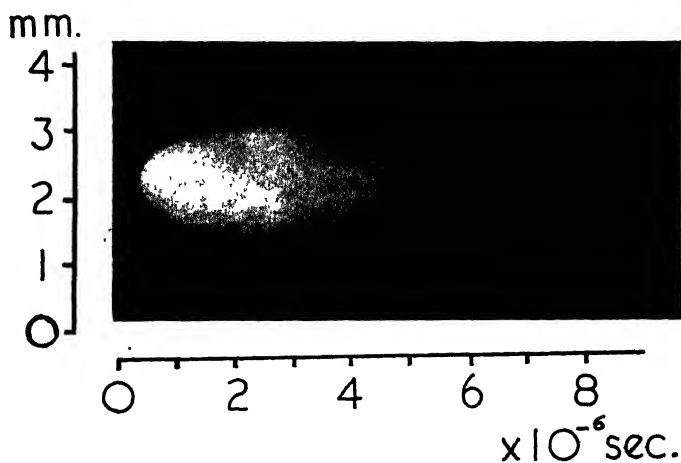


Figure 2. Rotating mirror photograph showing expansion of spark channel.

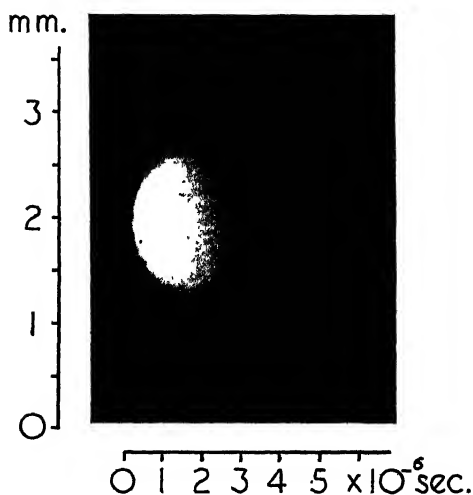


Figure 3. High resolution rotating mirror photograph showing initial expansion of spark channel.

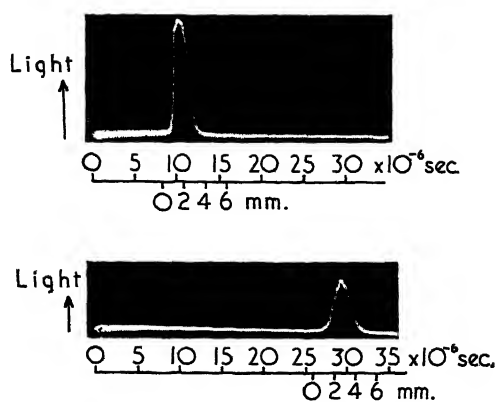


Figure 5. Typical oscillograms showing the distribution of light across spark channel.

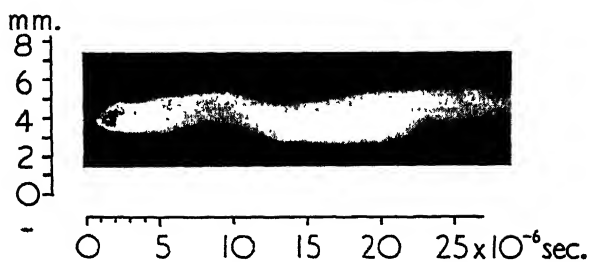


Figure 12. Rotating mirror camera record showing a typical hydrogen spark channel expansion.

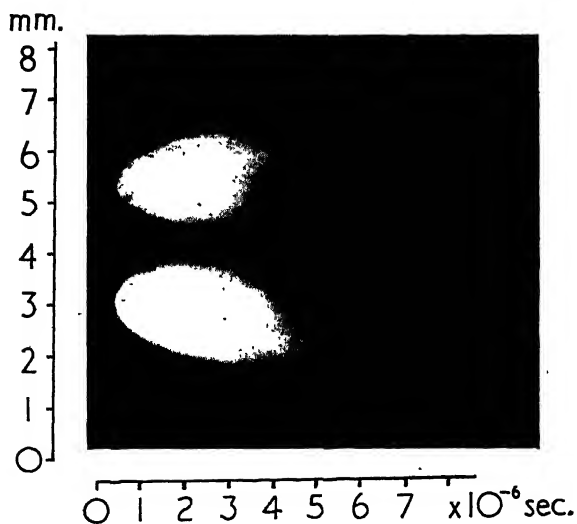


Figure 14. Rotating mirror camera photograph of a branched spark channel in nitrogen conducting a current of 500 amps.

The Expansion of Gaseous Spark Channels

By J. B. HIGHAM AND J. M. MEEK

Department of Electrical Engineering, University of Liverpool

MS. received 9th March 1950

ABSTRACT. A rotating mirror camera with a temporal resolution on the film of $1.5 \text{ mm}/\mu\text{sec.}$ has been used to determine the rate of expansion of long gaseous spark channels during their first $10 \mu\text{sec.}$ of growth. The spark channels conducted aperiodic impulse currents in the range of 60 to 500 amp. peak, which attained their peak values in about $1/4 \mu\text{sec.}$ and decayed to half-value in 10 or $28 \mu\text{sec.}$ Sparks have been investigated in air at atmospheric and reduced pressures, and in nitrogen, oxygen and hydrogen at atmospheric pressure. The measured areas of cross section of these spark channels bear a linear relation to the peak currents, except in hydrogen, and are little influenced by the rate of current decay in the range studied. Measurements have also been made with the rotating mirror scanning an image of the spark across a photomultiplier connected to a high speed oscillograph; by this method a direct record is obtained of the radial light distribution across the spark channels. The mechanism of spark channel expansion is discussed, and estimates are made of the average ion densities in spark channels.

§ 1. INTRODUCTION

UNTIL the studies by Flowers (1943), who recorded the luminous boundary of the spark channel by a rotating-film camera, there have been few investigations of the current densities occurring in spark channels, though a certain amount of information is available from earlier experiments by Beams (1930), Lawrence and Dunnington (1930) and Dunnington (1931), all of whom used a Kerr cell type of electro-optical shutter. Other measurements with rotating cameras have been made by Raiskij (1940, 1948), and more recently by Norinder and Karsten (1949).

In the present investigation (Meek 1947), a similar technique to that employed by Flowers (1943) has been adopted, but an improvement in time resolution has been obtained by the use of a rotating mirror instead of a moving film. This has enabled records to be made of the spark channel at earlier times subsequent to the formation of the channel.

As in the previous studies of spark channels it is assumed in the analysis of the results that the luminous boundary of the spark channel defines also the boundary of the region of ionization and therefore of the current flowing. This assumption has not been justified experimentally, but it is frequently adopted in electrical discharge measurements, and Suits (1935) considers that it is justifiable in the case of steady high-pressure arcs. Indirect evidence showing that the electrical boundary is greater than the visible boundary is given in the measurements of voltage gradient along sparks in tubes, but such evidence cannot be considered conclusive (Higham and Meek 1950). In the present records, although the light intensity varies across the spark channel, the luminous boundary is quite sharply defined and enables an effective diameter to be assigned to the spark. While the current density will also vary radially over the channel, an average effective current density may be defined as the total current divided by the effective cross-sectional area of the spark, and this definition has been used in the analysis of the results.

§ 2. MEASUREMENT TECHNIQUES

The present work was undertaken in parallel with the investigations on the voltage gradients in long spark channels (Higham and Meek 1950), and consequently the sparks used were of the same nature in the two cases. The point-point spark gaps were supplied from an 8-stage impulse generator of discharge capacitance $0.025 \mu\text{F.}$ through series resistors of 750 or 2,000 ohms. The currents attained their peak values in about $0.25 \mu\text{sec.}$, decayed to half-value in 10 or $28 \mu\text{sec.}$, and their peak amplitudes ranged from 60 to 500 amp.

2.1. Rotating-Mirror Camera

The optical arrangement of the rotating-mirror camera used for the measurements is shown diagrammatically in Figure 1. An image of the spark is formed by an object lens on an adjustable slit, of the type used in spectrographic work. The light passing through the slit from a short length of the spark is collected by a second lens and, after reflection by a rotating mirror, is brought to a focus on the photographic film. The resultant record then shows the

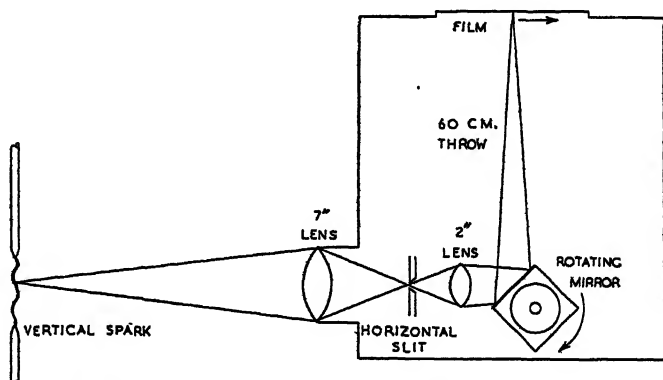


Figure 1. Schematic diagram of the rotating mirror camera.

photographed spark channel diameter as a function of time. The overall magnification from spark to film usually employed was 2 to 3 diameters, and, for a mirror rotational speed of 12,000 r.p.m., the time scale on the film is about $1.5 \text{ mm}/\mu\text{sec.}$, about four times faster than that used by Norinder and Karsten (1949). The rotational speed of the mirror was measured by a stroboscope.

A typical record showing the expansion of a spark channel is reproduced in Figure 2 (Plate I). For the purpose of analysis, the records were enlarged by about 10 diameters and tracings were prepared.

A necessary feature of the equipment is the arrangement for synchronizing the mirror position and the initiation of the spark so that when the spark occurs the mirror is correctly placed to reflect an image of the spark on to the film. For this purpose, a collimated light beam is reflected once each revolution of the mirror on to a photoelectric multiplier which is normally biased off. When the charging voltage of the impulse generator reaches a pre-determined value, the bias is automatically removed and the resultant pulse from the multiplier is amplified and fed to a blocking oscillator which provides a triggering spark (Husbands and Higham 1950) for the impulse generator. The instantaneous position of the mirror when the spark occurs can then be controlled so that a

record is obtained from each spark. With an 8-stage impulse generator the occurrence of the spark was controlled to within about $3\ \mu\text{sec.}$, represented by only 5 mm. on the 20 cm. length of exposed film. Full details of the circuits involved are published elsewhere (Prime and Saxe 1949).

Trial records were made with slits of various widths such as to give image widths in the plane of the film ranging from 5 mm. down to 0.3 mm., corresponding to exposure durations of 3.3 down to $0.2\ \mu\text{sec.}$ It was found that the channel diameters recorded with this range of slit widths were identical during the first few microseconds. The use of a narrow slit serves to improve the definition of the record for the initial stages of growth, as illustrated in Figure 3 (Plate I) which shows a record obtained with a slit image of 0.5 mm. Because of the short exposure involved, the channel is not recorded in its later stages, and it was found necessary to use a larger slit width. The slit usually employed gave an image in the plane of the film 2.5 mm. wide, corresponding to an exposure of $1.7\ \mu\text{sec.}$, and was found suitable for most of the sparks investigated. The record of Figure 2 was obtained with this slit. All the negatives show a clearly defined boundary to the light-emitting channel.

Both lenses in the camera were operated at $f/2.8$, and the 35 mm. film employed was Ilford H.P.3 fast panchromatic which was developed in MQ developer under standardized time and temperature conditions. This standardization enabled reasonable comparisons to be made between the photographed diameters of sparks for different currents and gas pressures. No attempt was made to compare actual light intensities, though this will be done in a future investigation after further work has been done in calibrating the film, as little is known on the response of photographic emulsions to the short duration exposures involved.

Analyses were made of about eight sparks for each set of conditions investigated, and it was found that the results were closely consistent throughout. The variation in measured diameters was rarely more than ± 0.15 mm. from one record to another.

2.2. Photomultiplier-Oscillograph Recording Equipment

A technique developed in the later stages of the research programme overcomes some of the difficulties inherent in photographic measurements. Some details of this technique have already been published (Saxe and Higham 1950). The rotating mirror camera is used without its resolving slit to scan an image of the spark across a slit in front of a photomultiplier, the spark and the photomultiplier slit being parallel to the axis of rotation of the mirror. The arrangement is shown schematically in Figure 4. By connecting the photomultiplier output through an amplifier (Prime and Saxe 1949) to a high speed oscillograph a record may thus be obtained of the light emission across a spark channel. Typical oscillograms are reproduced in Figure 5 (Plate II).

As in the photographic technique, the occurrence of the spark is synchronized with the mirror position, so that the image of the spark can be made to scan across the slit at any time relative to the initiation of the spark.

This technique, which was developed primarily for a future investigation to determine the distribution of light across a spark channel, also provides a check on the channel diameters as determined photographically in the present work.

A typical set of results is plotted in Figure 6 with the corresponding photographic results for comparison.

The photomultiplier method is more adaptable than the photographic method for investigating spark channels in their later stages when the light emission is weak, though photographic results can be obtained for these later

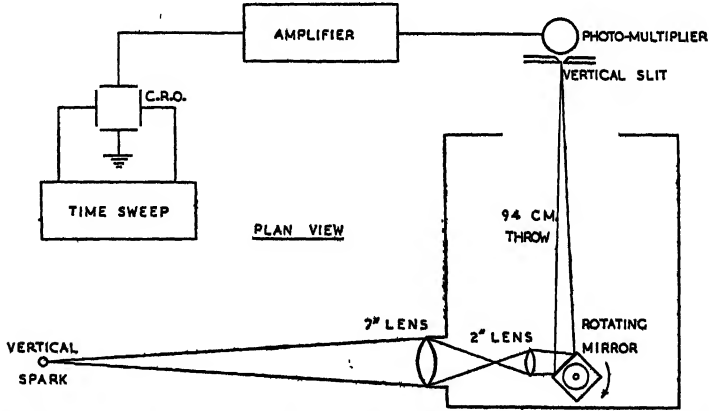


Figure 4. Schematic arrangement of the rotating mirror-photomultiplier-oscillograph recording equipment.

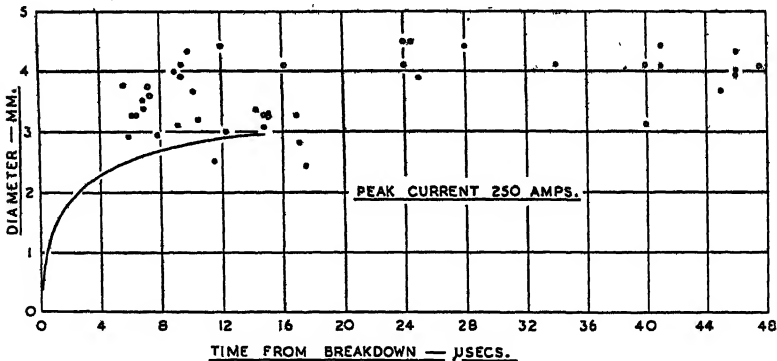


Figure 6. Spark channel diameters in air obtained with the rotating mirror-photomultiplier technique. The full curve represents results obtained with the photographic technique.

stages by using a wider resolving slit. However, the method is not so readily applicable to the initial stages of channel growth, when the diameter changes appreciably during the time of scan. A serious drawback to the multiplier technique is that it provides no definite means of indicating whether the spark is being scanned perpendicularly or obliquely to its axis. It was possible to see in the photographs whether the spark was perpendicular to the resolving slit, and if it was not perpendicular the record was discarded. This is probably the major factor causing most of the diameters determined by this method to be greater than the photographed diameters as shown in Figure 6. A further advantage of the photographic technique is that it provides a continuous record of diameter as a function of time, whereas only one cross section at a given time from breakdown is provided by each multiplier oscillogram.

§ 3. RESULTS

The results presented in §§ 3.1 to 3.3 refer to the spark channel diameter as photographed with the rotating mirror camera, though reference is made in § 3.1 to some results obtained by the photomultiplier technique.

3.1. Air at Atmospheric Pressure

Preliminary measurements made on sparks of various lengths from 1 to 40 cm. showed that the length had no influence on the channel diameter. Further, the diameter appears constant along the length of the channel except at positions close to the electrodes where the channel is affected by the presence of electrode vapour. Consequently it was decided to standardize on one gap length, of 10 cm., and to make measurements at a point in the channel about 1 cm. from one of the electrodes.

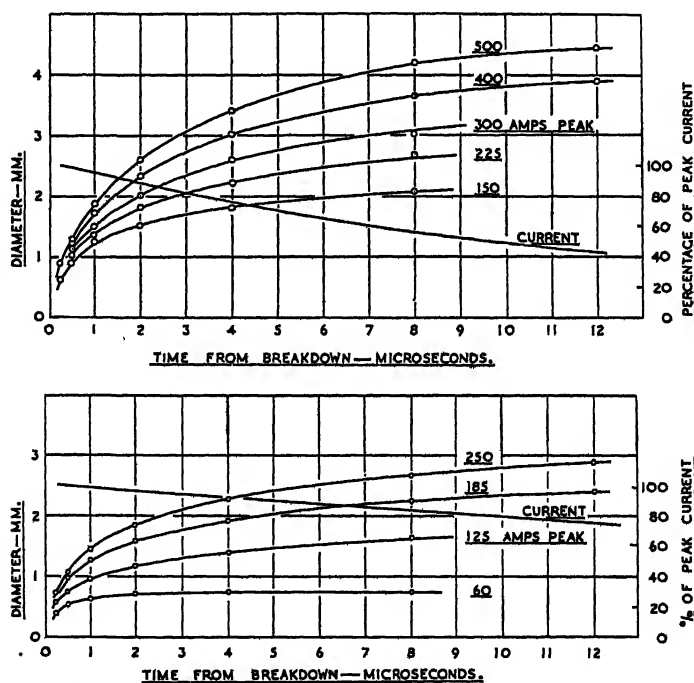


Figure 7. Photographically determined spark channel diameters in air as a function of time for various values of peak current and two rates of current decay.

A series of records was obtained for sparks across the 10 cm. gap for two current waveforms and for various values of peak current. The measured channel diameters are shown as a function of time in the two graphs of Figure 7 together with the corresponding current waveforms. The initial growth of the spark channel is very rapid, and, for a spark conducting a peak current of 500 amp. the diameter is increasing at a speed of about 9.3×10^4 cm. per sec. at $1 \mu\text{sec.}$ after the formation of the channel. At $8 \mu\text{sec.}$ from breakdown the rate of expansion has decreased to about 1.2×10^4 cm. per sec. With lower currents in the spark the initial rate of expansion is less rapid, being about 4.6×10^4 cm. per sec. at $1 \mu\text{sec.}$ after breakdown for a current of 150 amp.

However, after $10\text{ }\mu\text{sec.}$ the rates of expansion for the sparks carrying peak currents ranging between 150 amp. and 500 amp. are closely the same.

In Figure 8 the measured cross-sectional areas are plotted as a function of peak current, with time from breakdown as the parameter. For each rate of current decay employed the cross section is found to increase in a nearly linear manner with the peak current value. This implies that the current density in the sparks observed for a given rate of current decay is approximately constant at a given time from breakdown. With a peak current of 250 amp. the channel dimensions measured for the two rates of current decay are almost identical.

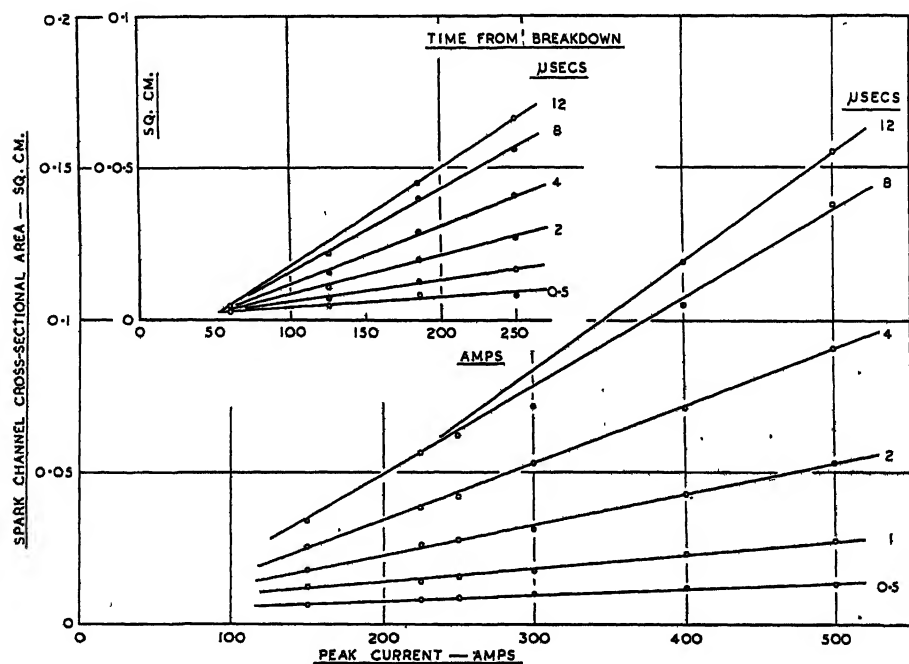


Figure 8. Photographically determined spark channel cross-sectional areas plotted as a function of peak current, with time from breakdown as parameter. The upper graph refers to currents decaying to half value in $28\text{ }\mu\text{sec.}$, and the lower graph to half value in $13\text{ }\mu\text{sec.}$

The lower rate of decay appears to be accompanied by smaller diameters below 250 amp. The curves therefore show that the peak amplitude of the current is a decisive factor in determining the diameter, while the rates of current decay for the two cases investigated are less important.

Some values determined for the average current densities in sparks in air at various times from breakdown are given in Table 1.

Table 1

Time from breakdown ($\mu\text{sec.}$)	0.5	1	2	4	8
Current	250 amp. peak decaying to half-value in $10\text{ }\mu\text{sec.}$				
Current density (k.amp/cm^2)	28.6	14.5	7.8	4.5	2.4
Current	250 amp. peak decaying to half-value in $28\text{ }\mu\text{sec.}$				
Current density (k.amp/cm^2)	29.0	15.0	8.9	5.6	3.7

The current densities in all the sparks recorded in air at a given time from breakdown are of the same order of magnitude as those given in Table 1, though higher densities were observed in the later stages of a channel conducting a peak

current of 60 amp., decaying to half-value in $28\ \mu\text{sec}$. This was the lowest peak current used in the experiments, and it is possible that the photographed image of the channel gives too small a diameter, because of the comparatively weak light emission, with a consequent apparent increase in the current density.

While the purpose of the investigations is primarily to determine the effective diameter of the spark channel, some measurements were also made, with the photomultiplier technique, of the variation of the light emission across spark channels in air. A number of oscillograms were analysed and the results are plotted in Figure 9 where the ordinate gives relative light intensity only,

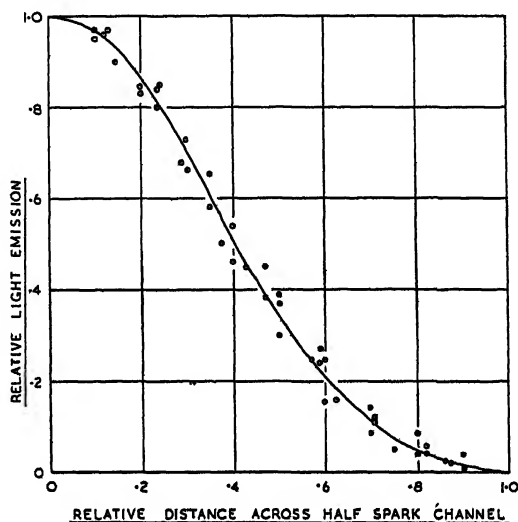


Figure 9. Results (to arbitrary scale) for the variation in light emission across a number of spark channels in air conducting currents ranging between 60 and 250 amps and at times from breakdown from 10 to $30\ \mu\text{sec}$.

with unit intensity at the centre of the channel. The oscillograms correspond to times between 10 and $30\ \mu\text{sec}$. after breakdown for spark channels carrying currents from 60 to 250 amp. It is clear that the distribution of light is closely the same for the range of conditions investigated.

3.2. Air at Reduced Pressure

Some of the measured spark channel diameters for air at reduced pressures are shown in Figure 10. It will be seen that down to 400 mm. Hg there is little change in the photographed diameter. Visual observation shows, however, that the channels produced at 100 and 200 mm. Hg are diffuse in nature, and this is borne out by the increased diameter recorded for the channel carrying a peak current of 500 amp. For channels carrying less current than this the light emitted is so weak that the records obtained are ill-defined for the later stages of growth, and these results have consequently been omitted from the graphs.

3.3. Nitrogen, Oxygen and Hydrogen at Atmospheric Pressure

Experiments were made with the 10 cm. spark gap enclosed in a Perspex cylinder 35 cm. in diameter and 90 cm. in length. Gases of commercial purity were used.

The results obtained for the various gases are shown graphically in Figure 11 with the corresponding results for air included for comparison. As for air, the important factor determining the channel diameter in the other gases studied is the peak current and not the rate of current decay. In nitrogen and oxygen the spark channels expand at about the same rate as that observed for air, and therefore the current densities in sparks in these three gases are closely similar.

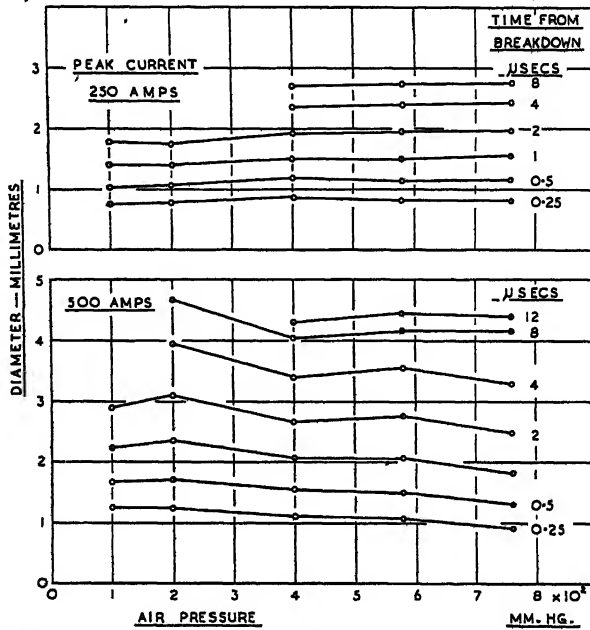


Figure 10. The effect of gas pressure on photographically determined spark channel diameters in air.

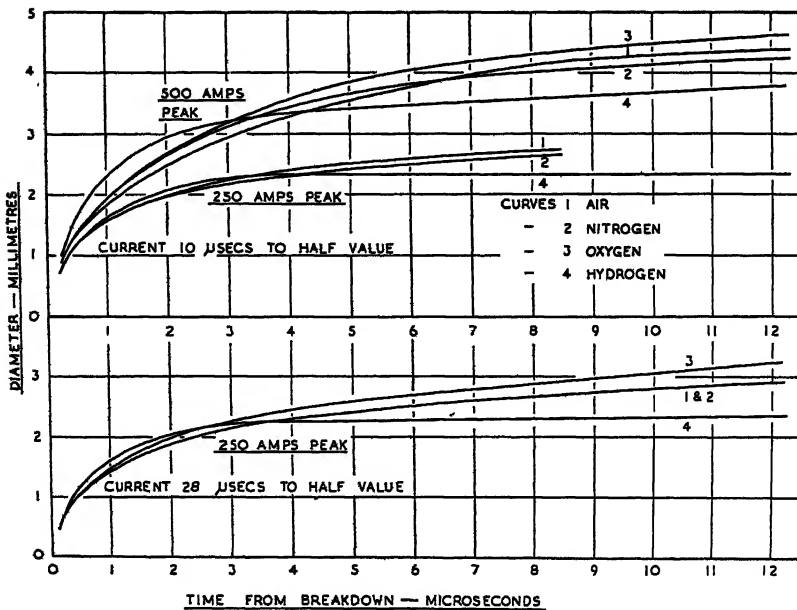


Figure 11. Photographed spark channel diameters in air, N₂, O₂ and H₂.

In hydrogen the expansion is more rapid in the early stages, but after about $5\mu\text{sec.}$ the channel reaches a nearly stable diameter. The result is that the current density is lower in hydrogen than in air during the first few microseconds after breakdown, but is higher in the later stages. Also, in contrast with air, the average current density in hydrogen appears to be greater for lower peak currents in the spark than for higher peak currents, as shown by Table 2, which gives the voltage gradients and current densities for sparks in hydrogen for currents decaying to half-values in $10\mu\text{sec.}$

Table 2

Peak current (amp.)	250		500	
Time from breakdown ($\mu\text{sec.}$)	2	8	2	8
Voltage gradient (v/cm.)	55	46.5	33.5	34.5
Current density (k.amp/cm ²)	6.95	3.6	6.55	2.9

Three other features of the hydrogen channel differentiating it from the channels in the other gases were noted, and are illustrated to some extent by the photograph of Figure 12 (Plate II).

(i) The light emission from the spark channel in hydrogen is practically constant until the current has decayed to a small fraction of its peak value, whereas for a spark in air under similar conditions of current, camera slit width etc., as illustrated by the photograph of Figure 2, the light emission has fallen considerably after about $10\mu\text{sec.}$ from breakdown when the current has decayed to half its peak amplitude.

(ii) The light emission is almost constant across the hydrogen spark channel, but attains a definite peak at the centre of spark channels in the other gases.

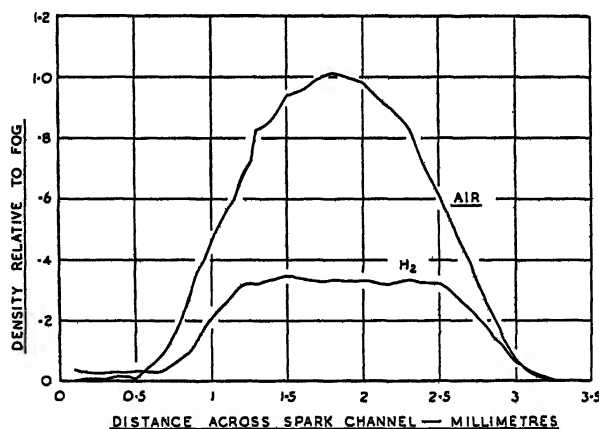


Figure 13. Photographic density measurements across rotating-mirror camera photographs of spark channels in air and hydrogen $3\mu\text{sec.}$ after breakdown. The current was the same in each case.

investigated. This is clearly shown by the photographic density measurements of Figure 13 which refer to typical spark channels in air and hydrogen at about $3\mu\text{sec.}$ after breakdown.

(iii) The photographs show that the long spark channels in hydrogen oscillate laterally. The amplitude of oscillation is approximately equal to the channel radius, and the frequency is of the order of 100kc/s. Most of the records exhibited this phenomenon but it is not possible to say definitely from the present

results whether the oscillation is in one plane or is a circular motion around the average axis of the spark. The latter seems the more likely, and would account for so many of the records showing the same amplitude of oscillation.

A rotating mirror camera photograph showing the expansion of a branched spark channel in nitrogen is given in Figure 14 (Plate II). 30% of the nitrogen records exhibited this effect, but it was rarely observed in the other gases. Branching on a larger scale, where the branch channels are well separated, was observed no more frequently in nitrogen than in the other gases.

A point of particular interest about these closely adjacent branches is the manner in which they expand, as all the records obtained show that the brightest portions of the channels decrease in cross section and move away from each other. Close examination of the original negatives reveals a dark layer of nearly constant thickness between the two branches. In the record shown there also appear two streamers of short duration close to the main discharge in its early stages of growth.

§4. DISCUSSION

As explained in the preceding paper (Higham and Meek 1950), a spark channel is initiated in a discharge gap when the space between the two electrodes has been bridged by a conducting plasma known as the leader stroke. The external circuit then commences to discharge through this plasma, producing the so-called main stroke (Schonland 1938, Allibone and Meek 1938). The current flowing in the main stroke depends on the external circuit but is generally many times higher than that in the leader stroke. The initial sectional area of the main stroke may then be assumed to be that of the leader stroke, but, because of the greater current and power expended in the main stroke than in the leader stroke, there is a rapid increase in the ionization within the channel with a consequent rise in temperature. The energy supplied to the spark channel is expended not only in producing the necessary ionization required to maintain the current in the channel but also in compensating the various losses caused by the diffusion of electrons and ions from the channel, by recombination processes, by light radiation and by the sound wave produced by the spark.

No complete theoretical analysis of the conditions within the spark channel is yet available, though discussions of various aspects of the problem are given in papers by Toepler (1925, 1926, 1927, 1929), Ollendorff (1933), Fucks and Bongartz (1943), Rompe and Weizel (1944), Craggs and Meek (1946), Craggs and Hopwood (1947), Norinder and Karsten (1949). The most detailed treatment is that by Ollendorff (1933) who has considered the particular case of the lightning channel, and, on the basis of the Saha equation and assumptions concerning diffusion, has deduced certain characteristics of the channel. However, the calculations are made for a discharge assumed to be stationary with respect to time, and no consideration is given to the rate of expansion of the channel.

Among the various factors governing the growth of the spark channel it is probable that both diffusion and recombination play important rôles, diffusion principally in determining the rate of expansion and recombination in influencing the density of ionization within the channel. Hopwood (unpublished work) has suggested that the liberation of photons from within the channel may also cause photo-ionization of the gas surrounding the channel and so extend the region of ionization, in this way causing an effective expansion of the channel.

In the initial stages of the spark the temperature of the electrons will be appreciably higher than that of the neutral atoms and molecules, but, as an

electron makes approximately 10^5 collisions per $\mu\text{sec.}$ with molecules in air at atmospheric pressure, the gas temperature, and therefore the gas pressure, within the channel will be greatly increased within a time of the order of $1\mu\text{sec.}$ Diffusion of electrons and positive ions will therefore occur from the ionized region, at a rate depending on the magnitude of the diffusion coefficients applicable to the particular conditions concerned. While electrons tend to diffuse at a more rapid rate than the positive ions, because of their higher speeds, the resultant excess positive space-charge field produced by positive ions left behind produces a field tending to retard the escaping electrons, and hence the expansion rate is probably governed by the so-called 'ambipolar' diffusion coefficient rather than by the diffusion coefficient for electrons alone. This expansion rate clearly relates to the electrical boundary of the channel and, as mentioned in §1, may or may not coincide with the recorded visible boundary.

Although it is probable that diffusion is the principal factor causing the radial expansion of the spark channel, calculations on this basis alone are not yet possible because of the present lack of knowledge of the distribution of ionization in the spark channel, and are further complicated by the fact that the supply of energy to the spark channel is continually changing. The diffusion rate may also be influenced by the high magnetic field surrounding the spark channel; for instance, the magnetic field strength at the circumference of a channel of diameter 1 mm. when carrying a current of 1,000 amp. is 4,000 oersteds. However, it has been pointed out by Craggs that the problem is greatly simplified in the case of the expansion of channels during the afterglow period immediately following the cessation of current flow, and calculations on this basis are being made in conjunction with experiments on sparks in hydrogen by Craig and Craggs (1950).

As the diffusion coefficients for positive ions in air, oxygen and nitrogen are roughly the same, the rates of expansion of spark channels in these gases may then be expected to be about equal. The diffusion coefficient in hydrogen is nearly five times that for the above gases, which probably accounts for the more rapid initial expansion observed experimentally.

The measured values for the current densities in air are in reasonable agreement with those of Flowers (1943) and Norinder and Karsten (1949), but a close comparison is not possible, partly because of the higher currents used by these investigators, and also because the times to peak current were appreciably greater than in the present work. Flowers concludes that the channel expands until the current density falls to a practically constant value of the order of 10^8 amp/cm.; the present results tend to show that a spark carrying unit function current of between 100 and 1,000 amp. would require about $50\mu\text{sec.}$ for the current density to fall to this value. Further expansion must occur, however, if the current is maintained, as eventually the discharge characteristics must become those of an arc, for which the current density is about 10 amp/cm² for currents of the order of 100 amp. (von Engel 1929).

Flowers does not differentiate in his measurements of expansion rates between the early stages and later stages of spark channel growth, and it appears that the resolution time of his camera was insufficient to record the changes occurring within the first one or two microseconds. However, Lawrence and Dunnington (1930) in their photographs of a spark carrying a current of 1,000 amp., record a cross section of about 10^{-3} cm² at a time $0.1\mu\text{sec.}$ after breakdown. This corresponds to a diameter of approximately 0.36 mm. and is in reasonable agreement with the value to be expected by extrapolation from the present results,

as given in Figure 7. In the studies by Raiskij (1940) a value of about 0.1 mm. is recorded for the spark diameter, but in this case the sparks were oscillatory and changes in diameter with time were not investigated. Later measurements by Raiskij (1948), concerned with the velocity of spread of the area of contact between sparks and electrodes, show that this velocity changes at the same rate as the velocity of expansion of the channel itself.

One of the interesting results of the present measurements is that the current density at a given time from breakdown is about the same for all the sparks observed, and is therefore roughly independent of the peak current, as shown by Figure 18. On the assumption that the current density remains the same in channels carrying much higher currents, the diameters of lightning channels may be estimated. For instance, with a current of 20,000 amp., which is the average magnitude of the current recorded in lightning discharges, the extrapolated value for the diameter is 2.7 cm. at a time of $10\mu\text{sec.}$ after breakdown, and is 3.2 cm. at $50\mu\text{sec.}$ Again, with a current of 100 kiloamp. the extrapolated diameter is 6.1 cm. at $10\mu\text{sec.}$ and 7.1 cm. at $50\mu\text{sec.}$ These extrapolated values are in rough agreement with the direct measurements made by Norinder and Karsten (1949), who find that the diameter of a spark channel carrying an oscillatory current, of quarter-period $12\mu\text{sec.}$ and peak amplitude 102 kiloamp., is 8 cm. at a time of $60\mu\text{sec.}$ after breakdown.

An estimate of the average electron concentration in the channel can be made from a knowledge of the average current density and the voltage gradient (Craggs and Meek 1946). On the assumption that the channel of cross section $A\text{ cm}^2$ contains uniform concentrations N_+ and N_- of positive ions and electrons per cm^3 , the current flowing in the channel is given by

$$I = (N_+v_+ + N_-v_-)Ae \text{ amp.},$$

where v_+ and v_- cm/sec. are the drift speeds of the positive ions and electrons respectively, and e is the electronic charge in coulombs. N_+ and N_- must be roughly equal, or otherwise large space charge fields would be set up. As $v_+ \ll v_-$, the term N_+v_+ may be neglected in comparison with N_-v_- , so that we may write

$$I = N_-v_-eA$$

or, if i denotes the current density, in amp/cm²,

$$i = N_-v_-e.$$

The value of the electron concentration N_- can then be computed from a knowledge of i and v_- . The value of i is given by the present experiments, while v_- depends on the voltage gradient in the channel, which is also known (Higham and Meek 1950). It may be assumed that in the short times considered (several microseconds) the gas density in the channel, and therefore the electron mean free path, is little changed from that of the surrounding gas, so that on this assumption the electron drift speed, corresponding to a particular voltage gradient in the spark channel, can be obtained by reference to electron mobility curves (Nielsen 1936). Values for N_- computed on this basis for spark channels in nitrogen at atmospheric pressure are given in Table 3. Clearly the values should be regarded as giving the order of magnitude only, in view of the approximation made by considering the current to be uniformly distributed over the channel cross section and the uncertainty regarding the applicability of electron mobility data to the different conditions obtaining in the spark channel. Again, mobility

data for molecular nitrogen have been used because, although practically 100% dissociation may be expected in the channel, no data are available for atomic nitrogen. The values for N_- are of the same order as those deduced in air and other gases in previous investigations (Craggs and Meek 1946, Craggs and Hopwood 1947).

Table 3

Current wave *	500 (10)			250 (10)			250 (28)		
Time from breakdown (μ sec.)	0.5	2	8	0.5	2	8	0.5	2	8
Channel diameter (cm.)	0.14	0.27	0.41	0.11	0.19	0.28	0.11	0.19	0.27
Current density (k.amp/cm ²)	32	7.9	2.2	29	7.8	2.4	29	8.9	3.7
Voltage gradient (v/cm.)	125	60	30	125	60	30	100	59	36
Electron drift velocity (km/sec.)	4	3.3	2.8	4	3.3	2.8	3.8	3.1	2.8
Electron density ($N \times 10^{-17}$ /cm ³)	5	1.5	0.5	4.5	1.5	0.5	4.8	1.8	0.8

* Figures give peak in amperes, decaying to half-value in () μ sec.

ACKNOWLEDGMENTS

This work was commenced in the high voltage laboratory of the Metropolitan-Vickers Electrical Company Limited and the authors wish to express their thanks, particularly to Mr. F. R. Perry, for the facilities provided and for the equipment they were allowed to retain.

The prosecution of the work has been greatly stimulated by the many discussions with Dr. J. D. Craggs, to whom the authors are indebted for advice in the writing of these two papers. The authors also thank their colleague, Dr. R. F. Saxe, for making available the oscillograph used in these investigations.

One of the authors (J. B. H.) makes grateful acknowledgment to Liverpool University for a Munitions Fellowship, and to the Institution of Electrical Engineers for Scholarships which enabled the work to be continued at Liverpool University.

REFERENCES

- ALLIBONE, T. E., and MEEK, J. M., 1938, *Proc. Roy. Soc. A*, **166**, 97, 169, 246.
 BEAMS, J. W., 1930, *Phys. Rev.*, **35**, 24.
 CRAGGS, J. D., and HOPWOOD, W., 1947, *Proc. Phys. Soc.*, **59**, 755.
 CRAGGS, J. D., and MEEK, J. M., 1946, *Proc. Roy. Soc. A*, **186**, 241.
 CRAIG, R., and CRAGGS, J. D., 1950, in preparation.
 DUNNINGTON, F. G., 1931, *Phys. Rev.*, **38**, 1535.
 VON ENGEL, A., 1929, *Z. tech. Phys.*, **10**, 505.
 FLOWERS, J. W., 1943, *Phys. Rev.*, **64**, 225.
 FUCKS, W., and BONGARTZ, H., 1943, *Z. Phys.*, **120**, 468.
 HIGHAM, J. B., and MEEK, J. M., 1950, *Proc. Phys. Soc. B*, **63**, 633.
 HUSBANDS, A. S., and HIGHAM, J. B., 1950, in preparation.
 LAWRENCE, E. O., and DUNNINGTON, F. G., 1930, *Phys. Rev.*, **35**, 396.
 MEEK, J. M., 1947, *Nature, Lond.*, **160**, 110.
 NIELSEN, R. A., 1936, *Phys. Rev.*, **50**, 950.
 NORINDER, H., and KARSTEN, O., 1949, *Ark. Mat. Astr. Fys.*, **36A**, Pt. 4.
 OLLENDORFF, F., 1933, *Arch. Elektrotech.*, **27**, 169.
 PRIME, H. A., and SAXE, R. F., 1949, *Proc. Instn. Elect. Engrs.*, **96**, Pt. II, 662.
 RAISKIJ, S. M., 1940, *J. Tech. Phys., U.S.S.R.*, **10**, 529; 1948, *J. Exp. Theor. Phys., U.S.S.R.*, **18**, 941.
 ROMPE, R., and WEIZEL, W., 1944, *Z. Phys.*, **122**, 636.
 SAXE, R. F., and HIGHAM, J. B., 1950, *Proc. Phys. Soc. B*, **63**, 370.
 SCHONLAND, B. F. J., 1938, *Proc. Roy. Soc. A*, **164**, 132.
 SUITS, C. G., 1935, *Physics*, **6**, 315.
 TOEPLER, M., 1925, *Arch. Elektrotech.*, **14**, 305; 1926, *Ibid.*, **17**, 61; 1927, *Ibid.*, **18**, 549; 1929, *Ibid.*, **21**, 433.

The Electrical Resistance of Liquid Gallium in the Neighbourhood of its Melting Point

By C. DODD

University College, London

Communicated by E. N. da C. Andrade; MS. received 17th May 1950

ABSTRACT. In order to determine whether any discontinuous change in electrical properties takes place when a liquid is supercooled, measurements have been made on the resistance of liquid gallium. No such discontinuity has been found, the resistance varying linearly with temperature over the range investigated.

The results yield values for the resistivity and temperature coefficient of resistance of liquid gallium.

§ 1. INTRODUCTION

RECENT experiments on supercooled liquids by Dodd and Hu Pak Mi (1949) have shown for several liquids a discontinuity in the viscosity-temperature curve occurring at the melting point as the liquid passes into the supercooled region. Since liquid gallium, a liquid metal (M.Pt. 29.9°C.), readily supercools, it is of interest to see whether any discontinuity occurs in any of its properties at the melting point. Viscosity and density measurements are at present being carried out in this laboratory, and the present paper describes the results of measurements of the electrical resistance of liquid gallium in the normal and in the supercooled state.

§ 2. PROCESSING OF GALLIUM

Gallium when melted, immediately acquires a surface coating of oxide which makes the metal wet glass, a condition unsuitable for accurate resistance measurements. About 40 gm. of gallium were melted in a quartz vessel and the oxide converted into the chlorides by addition of a little dilute hydrochloric acid. The vessel was exhausted through drying tubes and the excess acid absorbed in caustic soda tubes, leaving only gallium and a surface film of its chlorides. When the quartz vessel is maintained at red heat for two hours under a reduced pressure of 10^{-5} mm. Hg, the chlorides (boiling points 210°C. and 535°C.) boil off, leaving pure gallium with a shiny surface resembling clean mercury. This gallium does not now wet glass provided the vacuum is maintained. By tilting, the gallium was poured from the quartz vessel into the bulb of the resistometer joined to it by a cone and socket. The evacuated resistometer was then sealed off near this joint using a blow-pipe.

§ 3. APPARATUS

The resistometer (Figure 1) consists of a long (45 cm.) length of precision bore (1 mm.) capillary tubing in Pyrex with 1 cm. spheres blown at each end and connected to the reservoir bulb into which the gallium is initially introduced. Side tubes are fused to the spheres, and platinum strips sealed through the glass provide electrical contact between the gallium in the spheres and the mercury in the side tubes. The whole apparatus is mounted on a frame and immersed in a well-stirred water bath with the open ends of the side tubes just above the water level. The temperature of the bath was maintained constant to better than

0.01° c. and was measured with a mercury-in-glass thermometer calibrated to an accuracy within 0.02° c.

The electrical measurements were made on a Smith Difference Bridge in which the resistance of connecting leads is eliminated, so that resistances could be measured to an accuracy of about 0.00001 ohm.

§ 4. METHOD

The resistometer was first filled with liquid gallium and its resistance determined at a series of temperatures up to 20° c. above the melting point and also in the supercooled region. Readings taken at the same temperature for different fillings did not differ by more than 0.00004 ohm.

Since the measured resistance includes not only the resistance of the gallium in the capillary but also that of the mercury in the side tubes and of the platinum contacts, an estimation of these latter two resistances was obtained by measuring

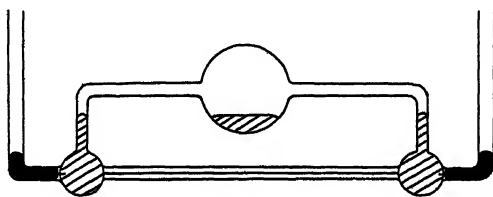


Figure 1.

the total resistance with the resistometer filled with pure mercury at 0° c. instead of gallium. The difference between this value and the value for the resistance of the mercury in the capillary tube, calculated from a knowledge of the resistivity of mercury and the dimensions of the capillary, gives the appropriate end correction at 0° c.

The uniformity and area of cross section of the tube were estimated by the usual mercury thread method and the length of the capillary was determined by a cathetometer.

§ 5. RESULTS

The measured resistance with the resistometer filled with Ga is plotted against temperature in Figure 2 over the range 0° to 50° c. It is seen that over this range the variation is linear, and moreover there is no indication whatever of any discontinuity occurring as the liquid passes through its melting point into the super-cooled region.

This confirms the work of Bridgman (1921) who made measurements on supercooled gallium at 0° c. and found the resistance to lie on a regular prolongation of the curve for the resistance above melting point. Guntz and Broniewski (1908) on the other hand found the resistance of the liquid to pass through a minimum and to increase again in the unstable region below the melting point.

§ 6. TEMPERATURE COEFFICIENT OF RESISTANCE

Neglecting the small resistance of the gallium in the end spheres, the measured resistance with the resistometer filled with gallium at temperature t ° c. is given approximately by

$$R_0 + \tau_0 + (R_0\alpha + \tau_0\beta)t$$

where R_0 is the resistance at 0° c. of the gallium actually in the capillary, τ_0 is

the resistance of the mercury in the side tubes also at 0°C. , and α and β are the temperature coefficients of gallium and mercury respectively. The slope of the line in Figure 2 is thus equal to $(R_0\alpha + r_0\beta)$, and the intercept on the resistance axis is $(R_0 + r_0)$.

At 0°C. Measured resistance when filled with mercury = 0.47752 ohm.

Calculated resistance of mercury in capillary = 0.47015 ohm.

Thus $r_0 = 0.00737\text{ ohm}$; $R_0 = 0.13606\text{ ohm}$; whence $\alpha = 1.089 \times 10^{-3}/\text{deg. C.}$

It may be noted that, like mercury, liquid gallium has a much lower temperature coefficient of resistance than most other metals for which the coefficient approximates more nearly to the theoretical value (3.66×10^{-3} per degree C.).

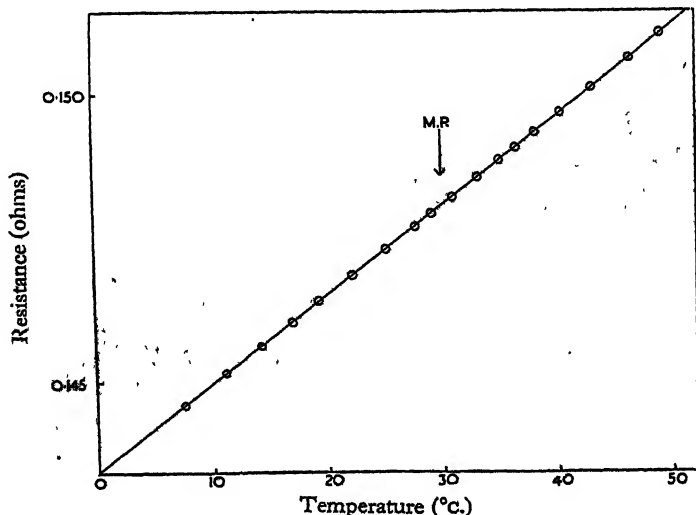


Figure 2.

§ 7. RESISTIVITY OF LIQUID GALLIUM AT 0°C.

The results obtained with gallium enable an estimate to be made of the resistivity of gallium at 0°C. , but in view of the uncertainty of the position of the ends of the capillary tube no great accuracy can be claimed. Since however the only published values for the resistivity differ widely, it is thought that even an approximate value might prove of use.

The area of cross section and the length of the capillary tube are $9.0438 \times 10^{-3}\text{ cm}^2$ and 45.20 cm. respectively, whence $\sigma_0 = 2.72_3 \times 10^{-5}\text{ ohm. cm.}$ This value may be compared with the value 2.592×10^{-5} found by Bridgman (1921) and 2.63×10^{-5} given by Guntz and Broniewski (1908) both these values being at the melting point.

ACKNOWLEDGMENTS

The author wishes to acknowledge his indebtedness to Dr. G. R. Davies of the Chemical Research Laboratory, Teddington, for kindly lending the supply of gallium, and to Professor E. N. da C. Andrade for the interest he continues to show in this and related work.

REFERENCES

- BRIDGMAN, P. W., 1921, *Proc. Amer. Acad. Sci.*, **56**, 109.
 DODD, C., and HU PAK MI, 1949, *Proc. Phys. Soc. B*, **62**, 454.
 GUNTZ, A., and BRONIEWSKI, W., 1908, *C.R. Acad. Sci., Paris*, **147**, 1476.

Hydrogen-filled Geiger Counters

By B. COLLINGE

George Holt Physics Laboratories, The University, Liverpool

Communicated by J. R. Holt; MS. received 27th February 1950

ABSTRACT. Permanent gas counters are discussed and a new quenching circuit is described. Quenching times of about $45 \mu\text{sec.}$ have been used and plateaux with slopes better than 0.05% per volt obtained. A formula for the correction of counting rates for losses caused by the counter dead-time has been confirmed within wide limits. It has been found possible to measure counting rates as high as 5×10^5 counts per minute with an accuracy closer than 1% .

§ 1. INTRODUCTION

SINCE the discovery by Trost (1937) that the presence of an organic vapour considerably improves the characteristics of counters, the use of counters containing a single gas has almost ceased. This is largely because counters containing an organic vapour have self-quenching properties, and a quenching circuit external to the counter is not required. In addition, these counters have a relatively short insensitive time, enabling them to be used at counting rates higher than was possible with externally quenched counters.

The self-quenching counter has, however, a number of disadvantages. The life is limited to about 10^8 or 10^9 counts and during this time its characteristics are gradually deteriorating as the organic vapour is broken down. Also, it is found that these counters are temperature sensitive. When a high degree of stability is required, externally quenched counters filled with a permanent gas have advantages. The dead-time of such counters is electronically controlled and can have a known, constant value, thus enabling an accurate correction to be applied for losses. It is shown in this paper that such counters can be made with better dead-times and plateau slopes than can be obtained with similar self-quenching counters.

Whether or not a counter contains an organic vapour, the discharge caused by the passage of an ionizing particle is terminated by the positive ions produced during the discharge. The positive ions form a sheath around the wire which reduces the field to a value below that at which multiple ionization can occur. In the self-quenching type of counter the organic vapour prevents the production of free electrons when the positive ions are collected at the cathode. Such electrons would produce a further discharge of the counter. Counters containing a simple gas only therefore require an external quenching circuit to ensure that the voltage across the counter is less than that required for a Geiger discharge when the positive ions are collected. The reduction of the wire potential caused by the circuit must not be too great, or the time for the positive ions to move across the counter may be unnecessarily long.

During the quenching time, that is the time for which the potential is reduced by the action of the circuit, the counter is operating in the proportional region. Hence an electron produced at the cathode causes a localized avalanche at the wire. The positive ions so produced will also be collected at the cathode and there is a chance of further electrons being produced. It may be necessary,

therefore, to quench for a period several times longer than the positive ion transit time, thus ensuring that no further electrons are produced in the counter after the wire potential has been restored to its full value. If γ is the probability of a positive ion producing an electron at the cathode and g is the gas amplification during the quenching time, then a condition for quenching to be effective is that γg shall be less than unity.

The time t for a positive ion to travel from anode to cathode in a Geiger counter has been shown by Stever (1942) to be given by

$$t = p \frac{(b^2 - a^2)}{2p_0KV} \log(b/a). \quad \dots\dots(1)$$

In the case of the counter used in the present investigations a , the radius of the wire, is equal to 0.1 mm., and b , the radius of the cylinder, is equal to 0.9 cm. Substituting these values in equation (1) a value of t of about 18.3 μ sec. is obtained for hydrogen at a pressure p of 28 cm.Hg, assuming a wire potential V of 2 kilovolts and a mobility K of 13.8 cm/sec/volt/cm. (Mitchell, quoted by Tyndall 1938); p_0 is the normal atmospheric pressure.

The quenching time necessary to ensure that there shall be no spurious counts must be greater than t ; the actual value will depend on γ . Values of γ , at the low energies possessed by positive ions at the cathode, have not been measured directly. The work of Penning (1930), however, indicates that values of γ for diatomic gases such as hydrogen and nitrogen are small compared with the values for the ions of helium, neon and argon. Cheyney (1917) has shown that γ may be very small for low energy hydrogen ions. This conclusion is confirmed by Jones (1939) who has calculated γ from data obtained from a discharge in hydrogen using copper electrodes. Jones found that γ for hydrogen was of the order of 10^{-6} for a value of x/p equal to 20. x is the field strength in volt/cm. and p is the pressure in millimetres of mercury.

The low value of γ and the high mobility of hydrogen ions suggest that counters filled with hydrogen might operate satisfactorily with quenching times considerably less than 100 μ sec. if a suitable circuit could be devised. In addition, hydrogen has other advantages as a counter gas: it is free from metastable states, it does not form negative ions by electron attachment, and counters filled with it produce larger pulses than when they are filled with other gases (Miller and Montgomery 1942).

§ 2. THE CIRCUIT

A number of circuits have been tried and found unsatisfactory at counting rates greater than a few thousand a minute. At these counting rates the counting action fails and a continuous discharge takes place. The reason for this is that the voltage on the counter returns to its full value and another discharge can occur before the circuit can be retriggered. Hence a discharge may take place which is not quenched and the chance that this may happen increases with the counting rate. An unquenched discharge will be followed by a spurious discharge which will in general have a smaller amplitude than a genuine pulse. This will be so if a discharge occurs before the field conditions in the counter are fully restored, either because the wire potential has not recovered, or because the spurious count occurred before all the positive ions had reached the cathode. If the circuit does not quench on this pulse then a second spurious discharge

will occur which may be of even smaller amplitude. Unless the circuit is sensitive enough to trigger on the small pulses which constitute a continuous discharge, the counter will cease to count; thus, not only must the circuit have adequate gain but its recovery time must be less than the transit time of the positive ions.

Further, it would be desirable for a quenching circuit to have a constant quenching time independent of the counting rate to enable a simple and accurate correction for counting losses to be applied.

The circuit shown in Figure 1 fulfils the necessary conditions. It consists of a separate quenching valve V_1 , the anode current of which is cut off by applying negative bias to the control grid. The counter wire is connected to the anode, which is normally at a positive potential of about 280 volts with respect to earth. The high-voltage supply for the counter is connected between the counter

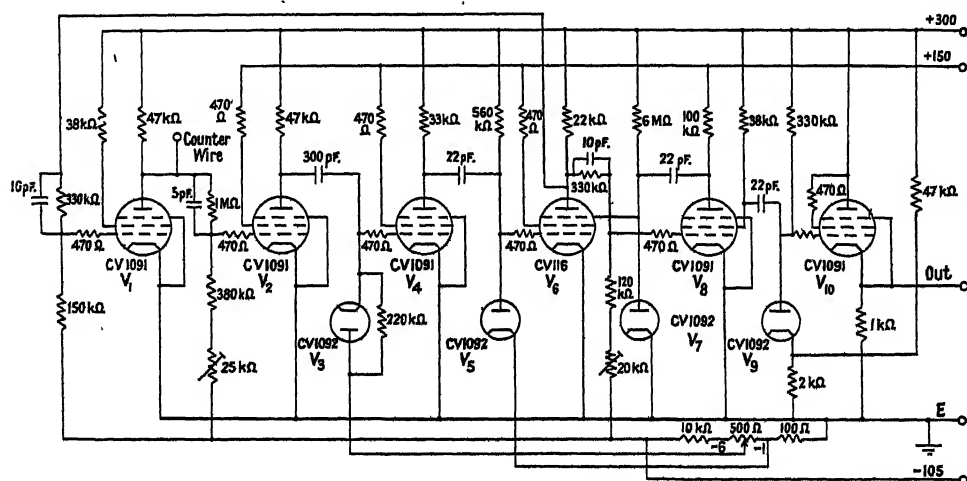


Figure 1. Counter quenching circuit.

cylinder and earth. After amplification, by V_2 and V_4 , the pulses from the counter trigger V_6 and V_8 , which are connected in a 'flip-flop' circuit; the period depends mainly on the circuit components. When V_1 conducts, the anode potential falls to within a few volts of the earth potential, causing the counter wire potential to be reduced by nearly 280 volts. The quenching time is conveniently adjusted by altering the value of the resistance between the suppressor grid of V_6 and H.T. The pre-set resistances compensate for component tolerances while the potentiometer in the grid circuit of V_4 is a gain control normally set for maximum sensitivity. V_9 and V_{10} have been included to provide a negative output pulse at low impedance and may not always be necessary.

§ 3. COUNTERS

A standard type of counter has been used in this work. It consists of a copper cathode, 12 cm. long and 1.8 cm. in diameter, mounted in a Pyrex glass tube and held in position by constrictions in the glass. A tungsten rod sealed into the glass wall and silver soldered to the cathode provides electrical contact. The tungsten wire is 0.2 mm. in diameter. Pyrex sleeves over the ends of the

wire protrude about 3 mm. into the cathode to define the sensitive volume of the counter.

Each counter was washed out with concentrated nitric acid and distilled water and baked out on a vacuum system at 375°C ., using a rotary pump and a diffusion pump for several hours. The copper cathode was reduced in hydrogen at 375° and then oxidized in air at the same temperature. In this way it was found possible to produce a uniform coat of oxide firmly attached to the cathode. The counter was baked out again for several hours on the pumps. After cooling it was filled with hydrogen free from oxygen and sealed off.

A pressure of 28 cm. of hydrogen was used, which gives an efficiency of roughly 99% for fast electrons (Korff 1946). Counting commenced when the potential between the counter cathode and earth was about 1,950 volts.

§ 4. MEASUREMENTS AND RESULTS

4.1. Quenching Time

The minimum value of the quenching time necessary to suppress spurious counts caused by electrons produced by positive ions at the cathode was determined as follows. The counter wire was connected to the anode of V_1 and a variable 2.2 kv. supply was connected between the cylinder and earth. The high-voltage supply was adjusted so that the counter was operating in the

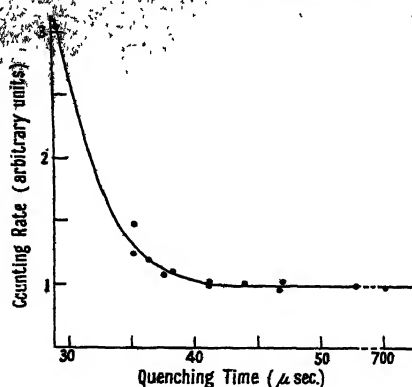


Figure 2. Counting rate as a function of the quenching time.

plateau region, and a series of readings of counting rate for various quenching times was obtained using a gamma-ray source to produce a suitable counting rate.

The results are shown in Figure 2. The counting rate remains constant as the quenching time is decreased from $600\mu\text{sec.}$ to about $40\mu\text{sec.}$ and it then increases rapidly at lower values.

4.2. Counter Characteristics

Curve (a) of Figure 3 is the characteristic curve of one of the standard counters; it was obtained with the circuit adjusted to have a quenching time of about $45\mu\text{sec.}$ The flat region of the plateau is 70 volts long and has a slope of 0.03 per cent per volt. The full length of the plateau is 170 volts. The circuit abruptly ceased to function at the upper end of the plateau.

Characteristic curves were obtained with other standard counters and gave results which were similar to this curve, but the slopes of the plateaux were not quite as good.

Curve (b) of Figure 3 is the characteristic curve of the same counter taken after an interval of three months. The starting voltage increased by about 40 volts although the slope did not change appreciably.

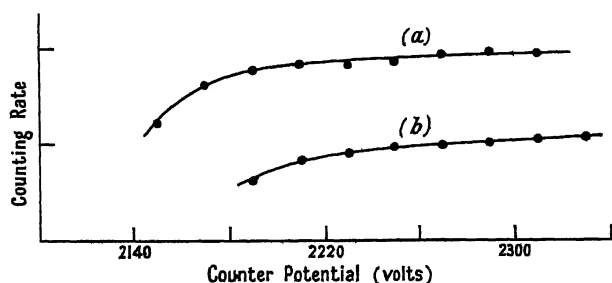


Figure 3. Characteristic curves of the standard counter.
(a) initially; (b) three months later.

A counter was made, similar to the standard counters, but with a reduced copper cathode which proved to be very photo-sensitive. It was noticed, immediately after running a continuous discharge, that the background counting rate was very high. Background counting rates of the order of a thousand per minute were observed immediately after switching the wire potential back to the operating value. The counting rate had not returned to the normal background rate after an hour. The standard counters with oxidized cathodes did not show this effect at all. It has not been found possible to suggest a satisfactory explanation.

4.3. Circuit and Counter Tests

The dead-time of the circuit was obtained using the paired source method. In this method two separate γ -ray sources are used and the values of the counting rates when the sources are presented to the counter, first separately and then together, are measured. The counter had the characteristic shown in curve (a) of Figure 3 and the readings were taken with the counter voltage adjusted to 2,280 volts. The value of the quenching time was determined from a calibrated cathode-ray oscillograph and is accurate to within about 3%.

Table 1

Quenching time ($\mu\text{sec.}$)	62.7	50.2	50.2
Dead-time ($\mu\text{sec.}$)	70.8	57.9	57.1
Difference ($\mu\text{sec.}$)	8.1	7.7	6.9

For each particular value of the quenching time ten separate determinations of the dead-time were made. The counting rate, obtained with two sources presented to the counter, varied from about 1,000 to 8,000 a second. The average value of the dead-time is shown in the second row of Table 1 for the quenching times shown in the top row. The dead-time was approximately constant within the limits of the experimental error over the range of counting rates used. The difference between these two figures is the recovery-time of the circuit and its average value is about 7.5 $\mu\text{sec.}$

The circuit and counters have also been checked using a half-gramme radium source and noting how the counting rate varied with its distance r from the counter. Each observed counting rate, N_r , was corrected for losses assuming a dead-time τ of $57 \mu\text{sec.}$, the value obtained in the previous experiment. The true counting rate N_t was obtained from the formula

$$N_t = \frac{N_r}{1 - N_r \tau} \quad \dots\dots(2)$$

In Table 2 the values of the product of r^2 and N_t are tabulated. The product should be constant. The bottom row of Table 2 is the fractional error, calculated from the mean, and indicates that the circuit and counters could be relied upon to better than 1.5% for losses below 20%. This is about the accuracy expected from the experiment. Measurements of the distance between the source and counter were unreliable to at least 2 cm. because of the finite size of source and counter, and furthermore no precautions were taken to reduce the effect of scattered γ -rays.

Table 2

Distance r between sources and counter (cm.)	241.6	281.6	321.6	351.6	371.6	401.6	451.6	501.6	551.6	601.6	631.6
Calculated true counting rate N_t (counts/sec.)	4645	3420	2617	2229	2082	1730	1379	1047	914	764	698
$r^2 N_t \times 10^6$	271	271	270.5	275	276	279	281	263	278	276	278
Percentage error in $r^2 N_t$	1.3	1.4	1.6	0	0.4	1.4	2.2	4	1.1	0.4	1.1

4.4. Correction for Counting Losses

It has been pointed out by Blackman and Michiels (1948) that equation (2) has not been satisfactorily confirmed experimentally. The apparatus described above enables an experiment to be performed to check the validity of the formula for large losses. A constant γ -ray source was used and the recorded counting rate measured for various values of the dead-time τ . The true value of the counting rate was calculated using equation (2) and the values of τ were obtained from an oscillograph which had been calibrated previously, using a pulse generator and a scale of 10^4 .

The results are shown graphically in Figure 4(a), (b) and (c). The calculated counting rate, obtained using equation (2), is plotted as a function of the losses expressed as a percentage of the true counting rate. The true counting rate was taken as equal to the calculated rate for a sufficiently small value of τ . Results are shown for three different positions of the source. The three curves have been combined in Figure 5. The ratio of the difference between the calculated counting rate and the true counting rate, expressed as a percentage of the true counting rate, is plotted along the vertical axis. The horizontal scale is the same as for the previous graph. The results lie on a smooth curve which should be a horizontal straight line. The extent of the deviations from this line indicates the accuracy to be expected from results obtained with the apparatus and corrected for losses using equation (2). The percentage error is less than 1% for a loss in the counting rate of 40%.

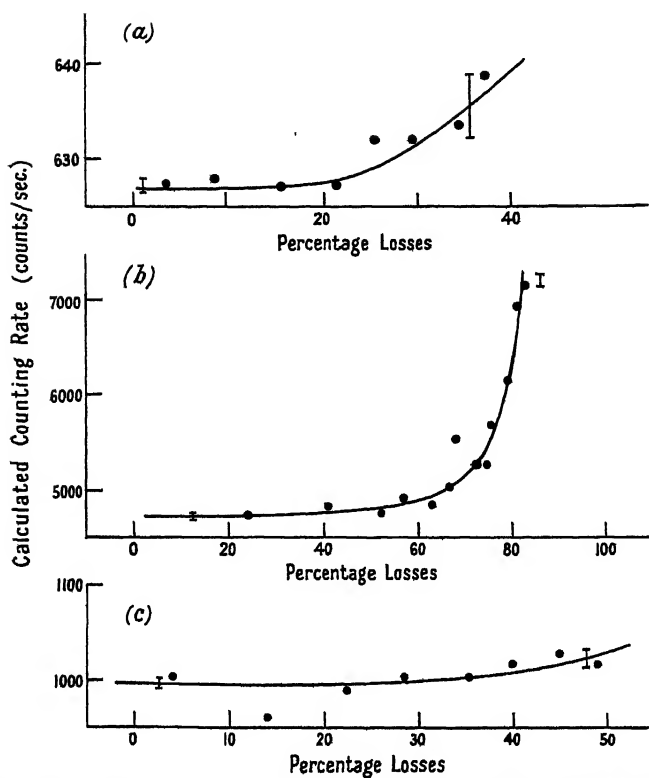


Figure 4. Calculated counting rate as a function of the percentage losses.

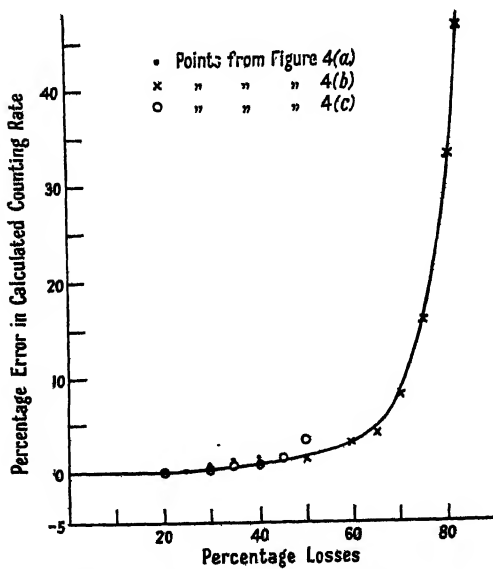


Figure 5. Percentage error in the calculated counting rate as a function of the percentage losses.

§ 5. DISCUSSION

5.1. *Ion Transit Times*

Mitchell (see Tyndall 1938, p. 70) has obtained values for the mobility of hydrogen positive ions in hydrogen for various values of x/p . His results show that for x/p less than 20 the mobility is constant and equal to 13.8 for moderately pure hydrogen. Hence equation (1) will be valid in regions where x/p is less than 20. It can be shown that x/p is greater than 20 for distances from the wire less than 1 mm. and that a positive ion will cross this region in less than 1 μ sec. Thus the figure of 18.3 μ sec. for the transit time of positive ions in the standard counter, calculated using equation (1), involves no serious error and should, therefore, agree with the observed value, if the disturbing effect of the positive ions on the field is neglected.

It is seen from Figure 2 that the minimum quenching time necessary to suppress spurious counts is about 41 μ sec. If the quenching time is reduced below 41 μ sec. then some spurious counts occur, indicating that the majority of the positive ions have reached the cathode in that time. The calculated time for one transit of the ions is 18.3 μ sec., that is, two transits take place in 36.6 μ sec. In the present case, therefore, quenching occurs when the quenching time is just greater than the time for two movements of the ions across the counter. The calculated and observed times are in good agreement.

Stever (1942) and others have found very poor agreement between the observed and calculated times for movements of positive ions in counters. Stever (1942) and Montgomery and Montgomery (1940) have suggested that the disagreement is due to variations of the mobility with x/p . Friedland and Krumbein (1948) have obtained better agreement by assuming a simple variation of the mobility with x/p . The results obtained here indicate that for hydrogen it is sufficient to assume a constant value for the mobility; this conclusion was reached by Hartog and Muller (1949) in the case of argon-alcohol mixtures.

5.2. *Measurement of γ*

The fact that only two transits of the ion sheath are required to ensure quenching indicates that γ is very small for hydrogen ions on oxidized copper. A rough calculation, based on the results shown in Figure 2, indicates that γ is less than 10^{-6} , in agreement with the work of Jones (1939). It may be possible to devise an experiment, similar in principle to the method used to obtain the results in Figure 2, in order to obtain values of γ . It would be necessary to measure both the number of ions in the ion sheath N_s and the gas amplification g during the time the counter is quenched.

5.3. *Length of Plateau*

It is seen from Figure 3 that the plateau is only 170 volts long, although the quenching voltage is 280 volts. The length of the plateau is determined, not by the quenching pulse, but by the potential across the counter for which the counter pulse is large enough to cause the valve V_2 to become non-conducting. The reason is as follows. Should a discharge occur which does not trigger the circuit and quench the counter, then the inevitable spurious count will not be amplified

by V_2 , and so quenching cannot occur. The counter then breaks down into a continuous discharge and no further counting is possible. This was confirmed by disconnecting the grid condenser of the valve V_4 and noting that V_2 became non-conducting when the counter voltage corresponded to the upper limit of the plateau.

5.4. Calculations of True Counting Rate

Lifschutz and Duffendack (1938) claim to have published an experimental verification of equation (2). Blackman and Michiels (1948) have pointed out that this work was not entirely satisfactory. In fact Lifschutz and Duffendack (1938) used a type I recorder which is re-excited by a further pulse occurring during the dead-time, whereas the formula only applies to type II recorders which have a constant dead-time. In the experiment described in this paper the apparatus had a constant dead-time and equation (2) is therefore applicable. From the results of Figure 5 it is concluded therefore that the formula for the correction of losses may be relied upon to better than 1% up to 40% losses.

The error caused by assuming that the dead-time of the circuit is equal to the quenching time may be neglected, because in this circuit the difference between these two times is a constant. A constant error in τ causes a constant error in the calculated counting rate, that is the curves of Figures 4 and 5 would be displaced vertically and the shape would be unchanged.

The shape of the curves can be explained by assuming that there is a small percentage error in τ which would occur if the duration of the pulse from the flip-flop circuit decreased as the on/off ratio of the circuit increases. A decrease of less than 10% in the quenching time is sufficient to explain the 50% error in the calculated counting rate at 80% losses.

No effort has been made to eliminate the small changes in the quenching time which, it is believed, cause the serious departures from the formula for losses greater than 40%. It seems unlikely that one would wish to correct for such large losses in practice.

5.5. Performance of Hydrogen-filled Counters

The quenching circuit used is somewhat more complicated than others described in the literature. It is believed to be superior, however, in that the quenching time can be precisely determined, thus enabling a simple and accurate correction for losses to be made up to high counting rates. It is implicit in this statement that the recovery time of the counter potential plus that of the circuit is short; the results shown in Table 1 show that it is of the order of 7.5 μ sec.

The performance of the circuit and counters can be summarized as follows:

- (i) adequate plateaux have been obtained using quenching times of 40 μ sec.;
- (ii) an accuracy to better than 1% can be expected at counting rates of 5×10^5 counts per minute when the losses are about 30%, assuming a dead-time of 50 μ sec.

From this it is evident that such counters would be useful when high counting rates are used over considerable periods of time and when it is necessary to have accurately reproducible conditions. The behaviour of a counter filled with a simple gas should be very reliable provided it is adequately outgassed. Such counters would be expected to have an almost unlimited counting life.

ACKNOWLEDGMENTS

The author is indebted to the Department of Scientific and Industrial Research for the grant which made the work possible and wishes to thank Dr. J. Rotblat and Dr. J. D. Craggs for many helpful discussions.

REFERENCES

- BLACKMAN, M., and MICHIELS, J. L., 1948, *Proc. Phys. Soc.*, **55**, 549.
 CHEYNEY, W. L., 1917, *Phys. Rev.*, **10**, 335.
 FRIEDLAND, S. S., and KRUMBEIN, A. D., 1948, *Bull. Amer. Phys. Soc.*, p. 51, April.
 DEN HARTOG, H., and MULLER, F. A., 1949, *Physica*, **15**, 789.
 JONES, F. L., 1939, *Phil. Mag.*, **28**, 192.
 KORFF, S. A., 1946, *Electron and Nuclear Counters* (New York: Van Nostrand).
 LIFSCHUTZ, H., and DUFFENDACK, O. S., 1938, *Phys. Rev.*, **54**, 714.
 MILLER, C. W., and MONTGOMERY, C. G., 1942, *Phys. Rev.*, **61**, 734.
 MONTGOMERY, C. D., and MONTGOMERY, D. D., 1940, *Phys. Rev.*, **57**, 1030.
 PENNING, F. M., 1930, *Proc. K. Ned. Akad. Wet.*, **33**, 841.
 STEVER, H. G., 1942, *Phys. Rev.*, **61**, 38.
 TROST, A., 1937, *Z. Phys.*, **105**, 329.
 TYNDALL, A. M., 1938, *The Mobility of Positive Ions in Gases* (Cambridge: University Press).

Cavitation produced by Ultrasonics

By B. E. NOLTINGK AND E. A. NEPPIRAS

Mullard Electronic Research Laboratory, Redhill, Surrey.

MS. received 9th February 1950

ABSTRACT. The problem of cavitation produced by ultrasonic vibration is examined theoretically. Equations are developed which describe the motion of a gas-filled cavitation bubble in a liquid medium subjected to alternating pressure; the case of an empty cavity is also considered. Information is obtained concerning the distribution of fluid pressures and velocities in the medium near the bubble surface during the motion. It is suggested that these theoretical conclusions may be used to show how the intensity of the various effects of cavitation will depend on ultrasonic frequency and intensity. In particular, it is predicted that all cavitation phenomena will diminish and finally disappear as the frequency is raised. The important part played by the nuclei from which the cavitation bubbles grow is emphasized.

§ 1. INTRODUCTION

IN recent years, a great deal of practical work has been done concerning possible applications of ultrasonic technique to industry. Many applications are known to depend on the production of cavitation in liquids, but no thorough examination has yet been made of what exactly occurs during ultrasonically induced cavitation.

The problem of erosion caused by cavitation was first encountered in connection with rotary hydraulic apparatus, such as ships' propellers and water turbines, and it was in order to obtain some measure of the liquid pressures capable of causing erosion that the early theoretical work on cavitation was attempted. Rayleigh (1917) examined theoretically the behaviour of an incompressible fluid in which he imagined a spherical void to be suddenly formed. More

recently, Beeching (1942) extended Rayleigh's analysis by taking into account surface tension effects, and the pressure of liquid vapour in the bubble. Silver (1942) introduced thermodynamic considerations, but several questionable assumptions in his treatment of the problem render his results of doubtful value.

The earlier experimental work was mainly confined to investigating the conditions under which cavitation would occur, but in 1948 Knapp and Hollander (1948) used high-speed cinematography to trace the life history of a cavitation bubble. They confirmed the existence of very large radial velocities and accelerations during the collapse period, for which the bubble motion was found to agree closely with that predicted by Rayleigh's treatment of the collapsing empty void. Quite recently, Plesset (1949) has developed an equation for the motion of a vapour-filled bubble in a changing pressure field and applied this to a complete analysis of Knapp and Hollander's experimental observations.

Usually, when cavitation is produced in hydraulic apparatus, the bubble will be at different points in the system at different stages of its life, and it is difficult, owing to the complex spatial variation of pressure, to be sure of the instantaneous pressure which is causing the bubble to alter in size. Plesset's calculations were based on the assumption that this pressure distribution was the same as that measured under non-cavitating conditions, while in Rayleigh's work, the pressure was simply taken to be constant. By contrast, in considering cavitation produced by ultrasonic waves, we need to be concerned with only one point in space, where the pressure changes are known and accurately controllable. This renders the problem amenable to more exact calculation.

§ 2. THEORY RELATING TO BUBBLE MOTION

(i) *Statement of Problem*

True cavitation (i.e. the formation of cavities or voids) in a pure liquid cannot occur until the liquid pressure has become sufficiently negative to overcome the forces of natural cohesion, and it is well known that degassed liquids can withstand very high negative pressures (Briggs, Johnson and Mason 1947, Harvey, McElroy and Whiteley 1947, Temperley 1947). But in a gas-filled liquid containing undissolved gas in the form of bubbles small enough to remain in suspension, these nuclei will expand or contract if the external pressure is made to change, and liquid will evaporate into the partial void produced. Although true voids are not present, this is the process generally referred to as cavitation when the pressure changes are sufficiently large and rapid, and it is the process we shall consider in the following theoretical discussion. As will appear more clearly below, no sharp distinction is to be expected between cavitating and non-cavitating conditions; the former is merely an extreme case.

The experimental evidence available (for example, Harvey *et al.* 1944) indicates that permanent gas nuclei are necessary for the onset of cavitation. These may, or may not, have come out of solution on solid nuclei. A third possibility exists when the liquid contains solid particles that are not wetted by the liquid; in this case, it is feasible that gas-free cavities may develop on what are essentially solid nuclei. Whatever their mode of origin, the bubbles will certainly gain vapour from the surrounding liquid as they grow, but since the pressure of vapour in equilibrium with a curved surface decreases exponentially with the curvature,

and we shall generally be concerned with very small bubbles, we may neglect the pressure of vapour in the bubble in comparison with the other forces controlling the motion.

In order to examine the problem theoretically, it is necessary to make further simplifying assumptions. The liquid is taken to be incompressible and the gas content of the bubble is assumed to be constant over its life cycle. We also assume that the applied ultrasonic pressure wave is exactly sinusoidal, although the very changes in the volume of the bubble under consideration must distort the pressure wave to some extent in its neighbourhood. Again, the assumption is made that the diameter of a bubble is always much less than a wavelength.

(ii) *Equations of Motion and their Solution*

The external liquid pressure at infinity, P , can then be written $P = (P_A - P_0 \sin \omega t)$ at time t , when an ultrasonic pressure wave of amplitude P_0 and frequency $\omega/2\pi$ is superposed on a pressure P_A . We wish to consider how a bubble will grow and collapse under this pressure. Let the bubble have an arbitrary radius R_0 at $t=0$ and contain gas at the equilibrium pressure $(P_A + 2S/R_0)$ where S is the surface tension of the liquid. The kinetic energy of the whole mass of liquid, density ρ , at this instant, $2\pi\rho R^3(dR/dt)^2$, can be equated to the algebraic sum of the work done by the surface tension, gas pressure and liquid pressure at infinity, giving as the energy equation :

$$\int_{R_0}^R \left\{ 4\pi R^2 \left[P_0 \sin \omega t - P_A + \left(P_A + \frac{2S}{R_0} \right) \frac{R_0^3}{R^3} \right] - 8\pi R S \right\} dR = 2\pi\rho R^3 \left(\frac{dR}{dt} \right)^2$$

if the gas changes are isothermal. The differential equation of motion is found by differentiation with respect to R , giving

$$2R \left[P_0 \sin \omega t - P_A + \left(P_A + \frac{2S}{R_0} \right) \frac{R_0^3}{R^3} \right] = 4S + 3\rho R \left(\frac{dR}{dt} \right)^2 + 2\rho R^2 \frac{d^2R}{dt^2} \quad \dots\dots(1)$$

If the nucleus consists of, not a gas sphere, but a spherical solid which is not wetted by the liquid, so that the cavity never contains any permanent gas, then $R=R_0$ when $P_0 \sin \omega t = (P_A + 2S/R_0)$ and the equation reduces to

$$2R[P_0 \sin \omega t - P_A] = 4S + 3\rho R \left(\frac{dR}{dt} \right)^2 + 2\rho R^2 \frac{d^2R}{dt^2} \quad \dots\dots(2)$$

In this case, it will be necessary for the liquid pressure to become appreciably negative before the void starts growing. These equations are insoluble mathematically, but a number of solutions were obtained on the differential analyser at the National Physical Laboratory. Examples are given in Figures 1 and 2, where $P = (P_A - P_0 \sin \omega t)$ is also plotted. It will be seen that the bubble radius increases to a maximum (coordinates t_m, R_m), and thereafter shrinks with increasing rapidity until the curve is, in fact, too steep to be traced out conveniently on the differential analyser. Such integrated curves were obtained for a range of values of the parameters P_0, ω and R_0 for a water medium at a pressure of one atmosphere ($\rho=1, S=80$ and $P_A=10^6$ c.g.s. units). It was found by trial that wide variations in the boundary values of radial velocity dR/dt at R_0 (which is arbitrary in equation (1)) produced insignificant changes in the contour of the

(t, R) curves; it was therefore taken as zero throughout. The values used for the various parameters are listed in the Table together with the corresponding values of t_m and R_m . It will be seen later that all the information required concerning the collapse of a bubble can be readily obtained once these coordinates are known and it is interesting to see how they depend on the primary variables. The

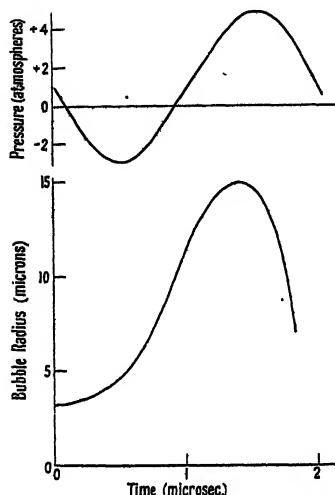


Figure 1. Radius-time curve for a gas-filled bubble.

$$P_0 = 4 \times 10^6; \omega = 3 \times 10^6; R_0 = 3.2 \times 10^{-4}.$$

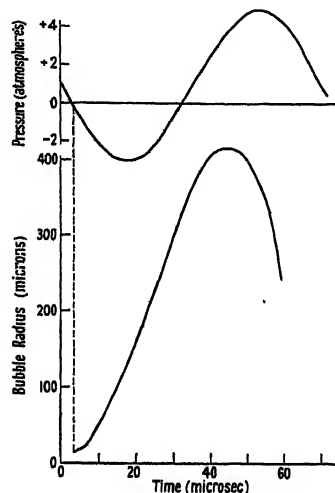


Figure 2. Radius-time curve for a void.

$$P_0 = 4 \times 10^6; \omega = 9 \times 10^6; R_0 = 16 \times 10^{-4}.$$

P_0 ($\times 10^6$ c.g.s. units)	ω	R_0 ($\times 10^{-4}$ cm.)	R_m ($\times 10^{-4}$ cm.)	t_m (μ sec.)
Obtained from solutions of equation (1)				
4	9×10^4	3.2	420	45.6
	3×10^5		128	13.7
	9×10^5		44	4.6
	1.5×10^6		28	2.8
	3×10^6	0.8	15	1.4
	9×10^6		6.7	0.5
	9×10^6		4.2	0.4
	3×10^7		1.7	0.15
	9×10^7		1.0	0.05
	9×10^4	80	496	47.3
		32	455	46.2
		16	438	46.0
2		3.2	420	45.6
		1.6	412	45.3
		0.8	407	45.3
		3.2	188	39.8
			318	43.3
			420	45.6
			579	48.0
			706	49.3
12			914	50.7
Obtained from solutions of equation (2)				
4	9×10^6	0.8	2.4	0.4
	9×10^4	16	420	46

curves of Figures 3, 4 and 5 show them plotted against P_0 , $1/\omega$ and R_0 respectively; Figure 4 is of particular interest in showing t_m and R_m to be roughly proportional to the period of the ultrasonic pressure wave over the range considered.

The very high pressures and radial velocities associated with cavitation occur during the collapse, when the walls of the bubble may be pictured as rushing inwards until they are repelled by the cushioning action of the gas in the bubble, which is by then extremely severely compressed. The total collapse time is usually a small fraction of a period of the ultrasonic vibration, so that no great errors will be introduced by regarding P as constant for its duration. The

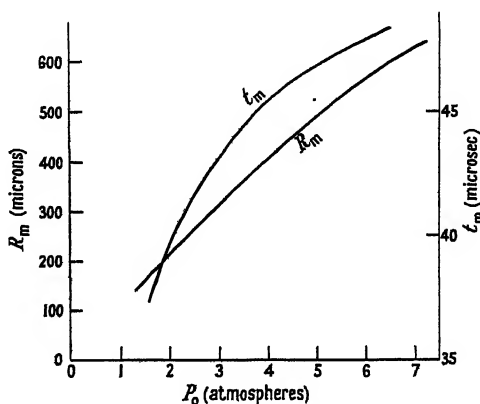


Figure 3. Variation of t_m and R_m with pressure amplitude, P_0 .
 $\omega = 9 \times 10^4$; $R_0 = 3.2 \times 10^{-4}$.

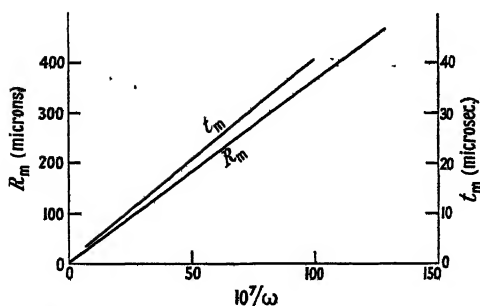


Figure 4. Variation of t_m and R_m with $1/\omega$.
 $P_0 = 4 \times 10^6$; $R_0 = 3.2 \times 10^{-4}$.

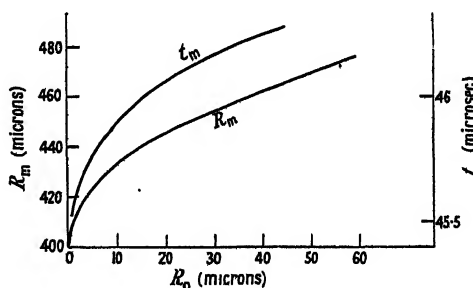


Figure 5. Variation of t_m and R_m with R_0 .
 $\omega = 9 \times 10^4$; $P_0 = 4 \times 10^6$.

motion of the bubble is then amenable to calculation; as shown in the Appendix, the limits of bubble radius and the maximum liquid kinetic energy and surface velocity of the bubble during the collapse can easily be deduced. This latter is given by V , where

$$V^2 = \frac{2P(\gamma-1)}{3\rho\gamma} \left[\frac{P(\gamma-1)}{Q\gamma} \right]^{1/(\gamma-1)}$$

if the gas changes are taken to be adiabatic, γ being the ratio of the specific heats for the gas and Q the pressure of gas in the bubble at its maximum radius. In many of the cases considered, V is greater than the velocity of sound in water, and since it is probable that fluid velocities do, in fact, often become supersonic during the final stages of the collapse, our theory, which assumes an incompressible

medium, must be thereby invalidated. However, it seems probable that deductions made from the theory, even when the liquid velocities are supersonic, can still give an indication of the relative importance of ω and P_0 in estimating the intensity of cavitation.

(iii) Oscillating Bubbles

We have so far only considered the behaviour of a cavitation bubble up to the instant when it collapses to its minimum size. This is obviously an unstable state, from which a series of complex oscillations would be expected to ensue. In the hydrodynamical treatment given, all forms of damping have been ignored. If these were taken into account, it is to be presumed that the amplitude of the oscillations would be found to fall off rapidly, and this is confirmed by the experimental work of Knapp and Hollander (1948).

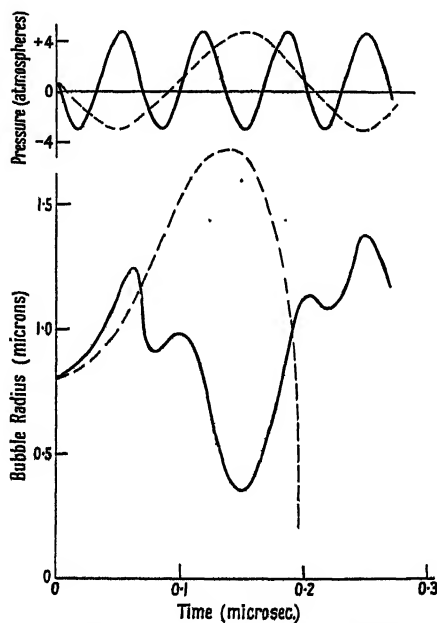


Figure 6. Radius-time curves for a gas-filled bubble at different frequencies.

$P_0 = 4 \times 10^6$; $R_0 = 0.8 \times 10^{-4}$; $\omega = 9 \times 10^7$ (full line), 3×10^7 (broken line).

While the events of the previous ultrasonic cycle may well be ignored if there has been time for a large number of damped bubble oscillations before the next cycle begins, matters will be different if the bubble period becomes comparable with the ultrasonic period. For $P_0 \ll P_A$, the resonant frequency of the bubble, n , is given by

$$(2\pi n)^2 = \frac{3\gamma(P_A + 2S/R_0)}{\rho R_0^3} \quad \dots\dots(3)$$

(Minneart 1933), the bubble motion being approximately simple harmonic. In this case, damping of the motion is due almost entirely to sound re-radiation and heat conduction (Saneyosi 1941, Meyer and Tamm 1939, Pfriem 1940). There is no doubt that these damping terms are important at ultrasonic frequencies even when, corresponding to $P_0 \ll P_A$, the amplitude of the motion is small.

If P_0 is comparable with, or greater than, P_A , resulting in large variations in the bubble radius, the oscillations will not be even approximately simple harmonic, and in this case, even if damping is ignored, an accurate expression for the resonant radius, corresponding to (3), cannot readily be deduced. However, equation (3) may still be used to give an order of magnitude for R_0 unless $P_0 \gg P_A$. For R_0 less than the resonant figure, in which case the collapse will be completed before the end of the pressure cycle, what may be described as true cavitation will occur; for R_0 greater than the resonant value, the oscillations may be very complex, but true cavitation should not occur. This prediction can be tested, in theory, by solving the equation of motion (1) for suitable high values of ω and P_0 and a value of R_0 greater than the resonant figure calculated from (3). At 15 Mc/s. and $P_A = 1$ atmosphere, this equation shows the resonant R_0 to be about 0.4 micron. The full line curve in Figure 6 gives a solution of equation (1) taken through four complete cycles of pressure at this frequency and $R_0 = 0.8$ micron. As can be seen, the oscillations are complex and nowhere do the radial velocities reach the high values characteristic of cavitating bubbles. The broken line curve refers to conditions identical with the above, except that the frequency has been reduced to 5 Mc/s.; as expected, it shows the steep collapse curve typical of cavitation. The inference is that for each impressed frequency there is a maximum to the size of bubbles that can give rise to cavitation. As the frequency is raised, this limit becomes smaller and smaller. It is thus apparent that, by sufficiently increasing the ultrasonic frequency, the other parameters remaining the same, the expected intensity of cavitation can be reduced to vanishing point.

§ 3. DEDUCTIONS FROM THEORY

For cavitating bubbles, the details of bubble motion and the conditions existing in the liquid medium surrounding the bubble during the collapse are dealt with in the Appendix; we now wish to show how the theoretical information obtained may be used to estimate the general intensity of cavitation action, and the way in which this varies with ω and P_0 . A difficulty which immediately presents itself is the arbitrary nature of the parameter R_0 ; we have little knowledge concerning practical limits to the possible values of this initial bubble radius. In theory, by taking it sufficiently small, whatever values we assign to ω and P_0 , we can obtain the conditions for producing cavitation pressures, etc., of almost any magnitude.* This means that the calculated magnitudes of liquid pressures, etc., obtained from the theory, may have little practical significance. However, we are primarily concerned with deciding how the intensity of cavitation will vary with ultrasonic intensity and frequency, and for that purpose it is perfectly legitimate to regard R_0 as constant, treating all our expressions as functions of ω and P_0 . The phrase 'intensity of cavitation' has an obvious general meaning, but before it can be used at all quantitatively, it must be given some more exact definition. It is probable that the function which best describes the intensity will depend on the phenomenon in connection with which the cavitation is being considered.

The bubble motion during collapse is decided entirely by the three parameters R_m , P and Q , the intensity of cavitation increasing with increasing R_m and P and

* It may be noted that there is a sharp distinction here dependent on whether the nuclei are taken as gaseous or solid. For the latter, if $R_0 < 2S/(P_0 - P_A)$ no cavitation can occur.

diminishing with increasing Q . P is governed in a simple way by ultrasonic amplitude, while Q for constant R_m depends only on R_0 , being proportional to $(P_A + 2S/R_0)R_0^3$. Since R_m has been shown to be roughly inversely proportional to ω , it follows that all cavitation effects will be expected to fall off with increasing frequency, the sharpness of this fall being dependent on the power of R_m which enters into the particular expression chosen to represent the effect being considered. We can, as previously suggested, expect cavitation to disappear completely in the region of

$$\omega = \frac{1}{2\pi R_0} \left[3\gamma \left(P_A + \frac{2S}{R_0} \right) \right]^{1/2}.$$

Considering the other end of the spectrum, it appears that cavitation intensity should increase indefinitely as the frequency is reduced. It is not yet clear how far this tendency may be limited by assumptions in the theory becoming invalid. One factor will certainly tend to prevent very large values of R_m being reached—namely, the reduction in the peak negative pressure of the ultrasonic wave occasioned by the very presence of large bubbles.

The curves of Figure 7, which are drawn from the information given in the Appendix, show how the amplitude of the liquid pressure wave following a collapsing cavitation bubble rises steeply as the bubble radius diminishes, and the wave velocity increases to very large values. The parameter Z used there

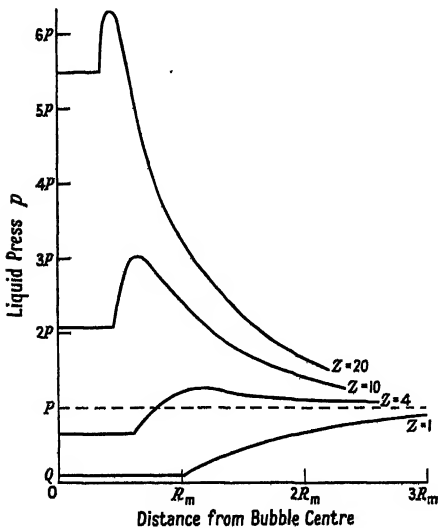


Figure 7. Pressure distribution in the fluid surrounding a collapsing, gas-filled bubble.

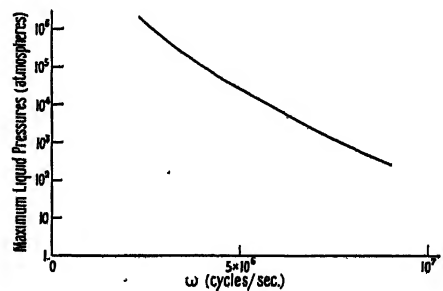


Figure 8. Variation with frequency of maximum fluid pressures during collapse. $P_0 = 4 \times 10^6$; $R_0 = 3.2 \times 10^{-4}$; adiabatic conditions.

is R_m^3/R^3 . The greatest fluid pressure during adiabatic collapse is $Q(P/3Q)^{3\gamma}$ and since $Q \propto R_m^{-3}$, this is $\propto R_m^{3(3\gamma-1)}$ or ω^{-9} approx., taking $\gamma = 4/3$ and $R_m \propto \omega^{-1}$. The graph of Figure 8 shows this maximum fluid pressure plotted against frequency for constant pressure amplitude and suitable R_0 . As can be seen, these maximum pressures are enormous, but high pressures of this order of magnitude exist for only a small fraction of the total collapse time and over a very small volume of fluid near the bubble surface. In some cases the absolute value of the pressure

set up in the medium may not be so significant as the excess above some threshold figure. If this threshold can be estimated, the information given in the Appendix can be used to show over what volume, and for how long, the fluid pressures remain effective. On this basis, a comparison might be made for varying frequencies and ultrasonic intensities. Integrated functions of fluid pressures obtained in this way may, in some cases, form a useful measure of the intensity of cavitation action, but it is probable that sometimes pressure gradients will be more important than the liquid pressures themselves. As shown in the Appendix, the hydrodynamical equations give the pressure gradient dp/dr directly as a function of distance, r , at any instant of the collapse, and, in such cases, this expression will be of value in deciding the effectiveness of cavitation action.

In the special case where gas-free bubbles grow from solid nuclei in suspension in the medium, an order of magnitude for the instantaneous pressure produced on the solid particles during collapse may be obtained by a simple calculation due to Rayleigh. He assumes the nuclei to be spherical, and abandons his assumption of incompressibility for the medium at the instant that the cavity wall comes into contact with the sphere. From that instant, he further assumes that the kinetic energy of each particle of fluid is changed to elastic energy of deformation of the same particle, as determined by the volume modulus of elasticity of the fluid, E . This gives $(P')^2/2E = (\rho/2)(dR/dt)^2$ where P' is the instantaneous pressure on the surface of the rigid sphere and dR/dt the velocity of the cavity wall on impact. The instantaneous pressure on the solid surface is therefore proportional to dR/dt . On Rayleigh's assumption of a completely empty cavity of arbitrary maximum radius, these pressures will obviously be greater the smaller the nuclei, becoming infinitely great as the nuclei approach zero radius. Our theory, applied to the gasless void growing and collapsing on a rigid sphere of known radius, involves no purely arbitrary assumptions (see equation (2)), although, as already mentioned, voids cannot develop unless $R_0 > 2S/(P_0 - P_A)$. The radius of the smallest nucleus on which voids can develop is therefore determined by P_0 and, since P_0 is limited by practical considerations, the collapse pressures are also limited.

Although, in the treatment given above, the gas changes for bubble growth are taken, for simplicity, to be isothermal and those for collapse strictly adiabatic, the true law obeyed by the gas content of the bubble will change continually during the bubble life, being more nearly adiabatic where the bubble volume changes more rapidly. The assumption of adiabatic conditions for the collapse period implies that the gas temperature will rise to very large values as the radius approaches its minimum. For a strictly adiabatic compression, the gas temperature at radius R is $T_0(R_m/R)^{3(\gamma-1)}$ and at minimum radius it is approximately $T_0(P/3Q)^{3(\gamma-1)}$, T_0 being the absolute temperature of the surrounding liquid. As an example, if $T_0 = 300^\circ \text{K}$, $P = 1$ atmosphere and $Q = 0.01$ atmosphere, the gas temperature at minimum radius would be about $10,000^\circ \text{K}$. Of course, in practice, the temperature of the bubble surface will be effectively T_0 , so that a steep temperature gradient must exist in the gas adjacent to it; how steep it is will depend on how nearly adiabatic the changes are. Since high temperatures are confined to the gas content, the liquid temperature remaining sensibly constant, it seems impossible that heating effects can play any important part in the destructive action of cavitation. On the other hand, the gas temperatures may rise to incandescence, so accounting for the luminescence that has been

observed under ultrasonic irradiation (Harvey *et al.* 1944). Various other effects are known to be produced by intense ultrasonic waves in liquids, including emulsification, dispersion, destruction of bacteria and other organisms, and several chemical actions. It may be possible to use the cavitation formulae that have been derived in analyses of these different phenomena.

§ 4. CONCLUSION

The theoretical treatment of the cavitation problem, contained in this paper, was undertaken to show how it may be possible, on the basis of pure theory, to predict the results of ultrasonic treatment under conditions of cavitation. Equations have been derived which describe the motion of a cavitating bubble, and the information obtained from them has been used to make a few broad deductions. Exact theoretical results are not to be looked for, but certain general laws can be laid down, in particular with regard to the effect of varying the ultrasonic frequency. For comparison with experiment, further investigations are required, and a fuller development of the theory. It seems likely that high speed cinematography with the techniques at present available could give definite checks on the calculations made, and could, in particular, show the importance of damping in relation to the bubble motion. It has previously been emphasized that all results will be markedly affected by the size and number of gas nuclei present in the treated liquid. While it is difficult, on the basis of information published to date to set any useful limits to the figures to be taken for these quantities, a number of experimental approaches suggest themselves which might throw more light on the subject.

APPENDIX

Assuming adiabatic gas changes, the energy equation for the collapse of a gas-filled bubble is

$$\frac{3p}{2} \left(\frac{dR}{dt} \right)^2 = P(Z-1) - \frac{Q(Z-Z^\gamma)}{(1-\gamma)} \quad \dots\dots(4)$$

where P , the external pressure, is taken as constant and $Z = (R_m/R)^3$, R_m being the maximum radius of the bubble at the start of the collapse. In this equation, surface tension forces are ignored, since they are quite unimportant during the collapse compared with the two opposing pressure terms. For stationary values of R , $dR/dt = 0$ giving $P(\gamma-1)(Z-1) = Q(Z^\gamma - Z)$ as the equation defining maximum and minimum bubble radii. This gives $Z=1$ or $R=R_m$ as one limit of oscillation. At the other limit, $R \ll R_m$, and $Z^{\gamma-1} = P(\gamma-1)/Q$ to a good approximation. The gas pressure corresponding to this minimum radius is

$$Q \left[\frac{P(\gamma-1)}{Q} \right]^{1/(\gamma-1)}$$

For stationary values of dR/dt , corresponding to points of inflection on the (R, t) curve, $(d/dt)(dR/dt) = 0$ or $(d/dR)(dR/dt) = 0$. This condition shows that, for maximum dR/dt ,

$$R = R_m \left[\frac{P(\gamma-1)}{Q\gamma} \right]^{1/3(1-\gamma)}$$

i.e. $R = \gamma^{1/3(\gamma-1)} R_{\min} \simeq 1.3 R_{\min}$ for air. The value of this maximum radial velocity is V where

$$V^2 = \frac{2P(\gamma-1)}{3\rho\gamma} \left[\frac{P(\gamma-1)}{Q\gamma} \right]^{1/(\gamma-1)}.$$

The total liquid kinetic energy when the bubble radius is R is

$$\frac{4\pi\rho R^3}{3} \left[P(Z-1) - Q \frac{(Z-Z^\gamma)}{(1-\gamma)} \right],$$

and this is a maximum where $(R/R_{\min})^3 = Q/P$.

To investigate the distribution of fluid pressure in the medium surrounding the collapsing bubble, we apply the general hydrodynamical equation of motion. If u is the fluid velocity at distance r from the centre of the bubble, the radial acceleration of the fluid at r is

$$\alpha_r = -\frac{du}{dt} = -\frac{\partial u}{\partial t} - u \frac{\partial u}{\partial r} = \frac{1}{\rho} \frac{dp}{dr},$$

p being the pressure at r . $\partial u/\partial t$ and $\partial u/\partial r$ can be obtained as functions of r and R from the energy equation and the continuity relation $R^2 dR/dt = r^2 dr/dt$. We obtain

$$\frac{dp}{dr} = \frac{R}{3r^2} \left[\frac{QZ^\gamma(3\gamma-4)}{(1-\gamma)} + \frac{QZ}{(1-\gamma)} - (Z-4)P \right] + \frac{4R^4}{3r^5} \left[P(Z-1) - Q \frac{(Z-Z^\gamma)}{(1-\gamma)} \right]$$

giving, on integration,

$$p - P = -\frac{R}{3r} \left[\frac{QZ^\gamma(3\gamma-4)}{(1-\gamma)} + \frac{QZ}{(1-\gamma)} - (Z-4)P \right] - \frac{R^4}{3r^4} \left[P(Z-1) - Q \frac{(Z-Z^\gamma)}{(1-\gamma)} \right]. \quad \dots\dots(5)$$

Plots of this function are shown in Figure 7 for various values of Z . To enable the curves to be drawn on a linear scale, the ratio P/Q is taken as low as 10 and Z not greater than 20. The maximum liquid pressure at any instant during the collapse occurs at a distance r_m from the bubble centre, obtained by solving the equation $dp/dr = 0$. The value of this maximum pressure is then obtainable from (5) by substituting $r = r_m$. It is easily seen that the greatest liquid pressure over the entire collapse occurs at minimum radius at the bubble surface and is equal, of course, to the gas pressure inside, QZ^γ . The pressure gradient at this point is $-QZ^\gamma/R_{\min}$ to a good approximation, assuming Z to be large. Parallel equations to (4) and (5) can be derived for the case where conditions are regarded as isothermal.

The equations relating to the collapse of a completely empty cavity can, of course, be obtained from the above by putting $Q = 0$. The energy equation reduces to

$$\frac{3\rho}{2} \left(\frac{dR}{dt} \right)^2 = P(Z-1).$$

This is the case considered in detail by Rayleigh (1917). The fluid velocities, pressures and pressure gradients approach infinite values as R approaches zero.

ACKNOWLEDGMENTS

This work was carried out under the auspices of the Mullard Radio Valve Company Ltd. The authors wish to thank the directors of the company for permission to publish, and Dr. C. F. Bareford, the Manager of the Electronic Research Laboratory for his encouragement in the project. Helpful discussions with Mr. J. G. L. Michel of the National Physical Laboratory are also gladly acknowledged.

REFERENCES

- BEECHING, R., 1942, *Trans. Instn. Engrs. Shipb. Scot.*, **85**, 273.
 BRIGGS, H., JOHNSON, J. B., and MASON, W. P., 1947, *J. Acoust. Soc. Amer.*, **19**, 664.
 HARVEY, E. N., 1939, *J. Amer. Chem. Soc.*, **61**, 2392.
 HARVEY, E. N., McELROY, W. D., and WHITELEY, A. H., 1947, *J. Appl. Phys.*, **18**, 162.
 HARVEY, E. N., WHITELEY, A. H., McELROY, W. D., PEARSE, D. C., and BARNES, D. K., 1944, *J. Cell. Comp. Physiol.*, **24**, 23.
 KNAPP, R. T., and HOLLANDER, A., 1948, *Trans. Amer. Soc. Mech. Engrs.*, **70**, 419.
 MEYER, E., and TAMM, K., 1939, *Akust. Z.*, **4**, 145.
 MINNEART, M., 1933, *Phil. Mag.*, **16**, 235.
 PFRIEM, H., 1940, *Akust. Z.*, **5**, 202.
 PLESSET, M. S., 1949, *J. Appl. Mech.*, **16**, 277.
 RAYLEIGH, LORD, 1917, *Phil. Mag.*, **34**, 94.
 SANEYOSI, Z., 1941, *Electrotech. J., Tokyo*, **5**, 49.
 SILVER, R. S., 1942, *Engineering*, **154**, 501.
 TEMPERLEY, H. N. V., 1947, *Proc. Phys. Soc.*, **59**, 199.

The Colour Temperature of Light Sources

By H. G. W. HARDING

National Physical Laboratory, Teddington, Middlesex

*Lecture given to the Colour Group of the Physical Society on 4th June 1947;
 revised MS. received 8th February 1950*

ABSTRACT. The definition of colour temperature is explained and the approximations of the spectral distributions of energy of lamps to those of black-body radiators are indicated. The value of colour-temperature measurements to the colorimetrist is outlined and difficulties in making the measurements are dealt with. The use of photoelectric devices for colour-temperature measurements are mentioned.

§ 1. DEFINITIONS

THE International Lighting Vocabulary definition of colour temperature, with its extended meaning, states that if the standard Commission Internationale de l'Éclairage (C.I.E.) observer can colour-match the radiation from the light source with that from a full radiator (black-body radiator), then the colour temperature to be assigned to the light source is given by the temperature of the full radiator when a colour match is obtained. The light source may be a full radiator, when for a colour match there will be a spectral energy match; it may be a grey body, when for a colour match there will again be a spectral energy match, or it may be a selective radiator, when for a colour match there will probably not be an energy match.

Before going any further a full radiator, grey body and selective radiator should perhaps be defined.

A full radiator completely absorbs all radiations incident on it whatever their nature or direction. The spectral distribution of energy of the radiation emitted by it is determined solely by its temperature, and of all incandescent bodies, at the same temperature, the full radiator emits the maximum total radiation.

A grey body emits radiation the spectral distribution of energy of which bears a constant ratio at all wavelengths to that of a full radiator at the same temperature, that is, for a given temperature the colour is the same as that of the full radiator, but it is not so bright.

A selective radiator emits radiation the spectral distribution of energy of which does not bear a constant ratio at all wavelengths to that of a full radiator at the same temperature. The spectral distribution of energy of a selective radiator can, however, be almost the same as that of a full radiator at another temperature; tungsten is an example of this.

§ 2. INCANDESCENT FILAMENTS

As long ago as 1914 Paterson and Dudding in their experiments referred to the Colour Identity Temperature of platinum, tungsten and carbon lamps, and it is to light sources of this nature, that have spectral emission curves almost identical with those of grey bodies, that the term colour temperature seems to have been first applied. Evidence for believing that tungsten and carbon lamps have spectral energy distributions that are approximately the same as those of full radiators is outlined by the following experiments. Worthing (1917) showed that heated tungsten is not a grey body.

A tube of tungsten with small radial holes in it $\frac{1}{10}$ to $\frac{1}{5}$ mm. in diameter was heated and at $2,600^{\circ}\text{K}$. the inside and outside temperatures were within about 3°K . of each other. It had the appearance shown in Figure 1 (Plate). The broad horizontal patch represents the tube and the small bright spot the hole in it. Owing to internal reflections, black-body radiation is emitted from the hole so that it is brighter than the rest of the tube. The vertical thin line is the hot wire of the pyrometer, the brightness of which was varied to match first the tube then the spot. Measurements were made for blue light of wavelength 0.467μ and red light of wavelength 0.665μ . Typical values for the ratios of the brightness of the radiation from the tungsten surface to that from the hole when the tungsten is at $2,600^{\circ}\text{K}$. are 0.46 for the blue light and 0.42 for the red. These values are, of course the emissivities and if the tungsten had been a true grey body the emissivities would have had the same value.

Forsythe (1923) followed on these measurements using light of three wavelengths. In effect, the temperature of the glowing tungsten was altered until the red to blue ratio was the same as that of a full radiator, then, without altering the temperatures of either the tungsten or the full radiator, the relative brightnesses for green light were measured. The ratio for the green light was found to be almost the same as it was for the red and blue. For carbon no difference was measurable. Tungsten, platinum and tantalum emitted about 1% more green light than a full radiator at $2,600^{\circ}\text{K}$. and osmium slightly less light. These experiments show that tungsten is not a true grey body in the sense that if it has the same temperature as a full radiator the spectral emission curve is the same shape. They show, however, that if the temperature is altered so that the colour

is the same as that of a full radiator the curves, as defined for light of three wavelengths, do become almost identical in shape. Confirmatory evidence of this for light of other wavelengths is given in Figure 2.

This experimental information indicates that the lamp filaments generally used emit radiation with spectral energy distribution curves that are practically the same shape as those of a full radiator at the same colour temperature. Provided that the temperatures of the filaments are reasonably uniform and the lamp bulbs are colourless, lamps made with these filaments can be used as substitutes for full radiators.

Now, if it is known that a tungsten filament lamp has a colour temperature θ , that is, it colour-matches a black-body furnace at a temperature θ , what can be deduced from this knowledge?

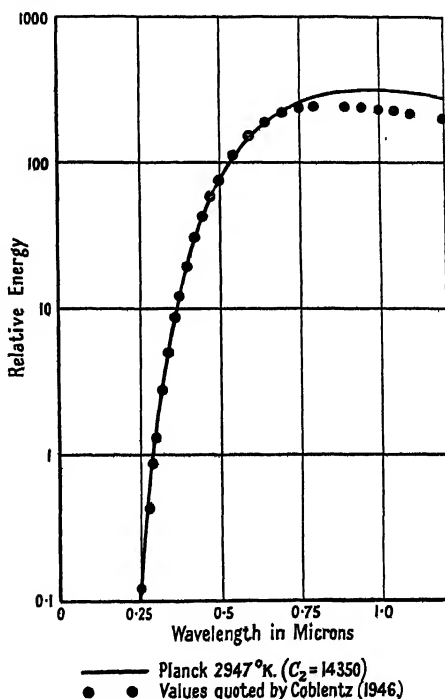


Figure 2. Spectral distribution of energy from tungsten ribbon in fused quartz bulb. After Coblenz, 1946. Reproduced by kind permission of the American Astronomical Society.

First the relative spectral energy distribution can be calculated by substituting the colour temperature for θ in Planck's formula

$$E_{\lambda\theta} = \frac{C_1 \lambda^{-5}}{\exp(C_2/\lambda\theta) - 1} \quad \dots\dots(1)$$

The only unknown for the evaluation of the relative distribution is C_2 , which for colorimetric purposes has generally been taken as being equal to 14,350 micron degrees. The calculation is laborious as six-figure accuracy is generally required for the preparation of standard tables. Makowski (1949) of the Admiralty Research Laboratory, Teddington, has made a slide rule which gives between two and three figure accuracy. This rule is exceptionally useful for experimental work as it gives a great deal of information in addition to the relative energies.

Having the relative energy distribution, calculated by substituting the colour temperature for θ in Planck's formula, the chromaticity of a lamp can be calculated from this energy distribution and the standard C.I.E. distribution coefficients for the equal energy spectrum (Harding and Sisson 1947), and this chromaticity can then be plotted on the chromaticity chart. This procedure is adopted in colorimetry for tungsten filament lamps and these lamps provide a wide range of light sources of known chromaticities and spectral emissions.

§ 3. C_2 IN PLANCK'S FORMULA

With reference to the value of C_2 in Planck's formula it should be mentioned that in 1948 the Comité International des Poids et Mesures (C.I.P.M.) adopted a modification of the international high temperature scale that involved a change in the accepted value of C_2 . This change affects the colour-temperature scale and through this the properties of standard illuminants for colorimetry defined by the C.I.E. in 1931. Methods of dealing with the problems created are at present under consideration by the appropriate bodies here and elsewhere.

The fact that the relative spectral distributions of energy depend only on the ratio C_2/θ helps to a limited extent in estimating the effects of changes in the value assigned to C_2 . If an energy distribution has been calculated for a temperature θ with a value C_2 , then if C_2 is altered to C_2' this same distribution will serve for another temperature θ' given by

$$C_2/\theta = C_2'/\theta'. \quad \dots\dots(2)$$

§ 4. ESTIMATION AND MEASUREMENT OF COLOUR TEMPERATURES

4.1. *Direct Method*

The measurement of colour temperature of the lower temperature range is straightforward and is illustrated by Figure 3 (Buckley, Collier and Brookes 1924).

A full-radiator furnace, a photometer and a comparison lamp are set up as shown. The lamp is colour-matched with the furnace and then a standard lamp is placed on the furnace side of the photometer and is matched against the comparison lamp. The temperature of the furnace is measured with the pyrometer which views it through the prism. The basis for the furnace temperature scale (International Temperature Scale C.I.P.M. 1948) is the melting point of gold, Planck's radiation formula with C_2 equal to 14,380 micron degrees and the absolute temperature as -273.15°C . The gold melting point, determined by gas thermometry, has the internationally fixed value of $1,063^\circ\text{C}$. or $1,336.15^\circ\text{K}$. The pyrometer measurements are brightness-temperature measurements and are made with a red filter over the eyepiece so that the pyrometer only measures the brightness of the furnace compared with the brightness of gold at its melting point, for red light. Furnace measurements were made up to $2,400^\circ\text{K}$. for the present N.P.L. scale, but higher temperatures have been used both here and abroad. Tungsten filament lamps will operate at temperatures up to about $3,640^\circ\text{K}$., when the filaments melt. The extension of the scale up to this temperature using real furnaces is not possible at present and it is not desirable that standard lamps should be calibrated at the higher temperatures because their life will be so short. Indirect methods have therefore to be used in order that the standard lamps shall operate at comparatively low temperatures when a reasonable life can be expected from them.

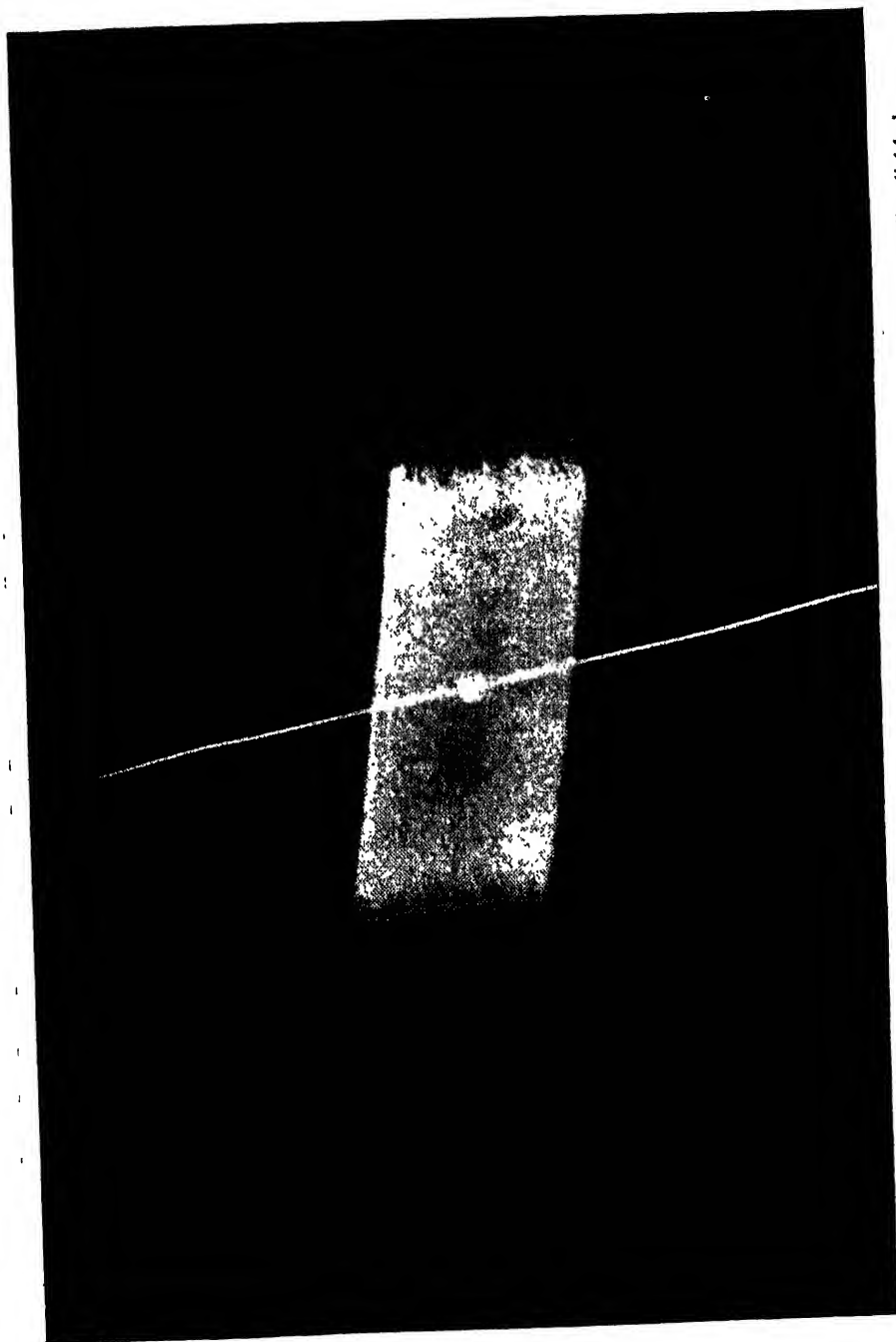


Figure 1. Photograph showing tungsten tube black-body with pyrometer filament projected against small radial hole.
Reproduced by kind permission of the International Commission on Illumination, New York.

1. 1

2. 1

3. 1

4. 1

5. 1

6. 1

7. 1

8. 1

9. 1

10. 1

11. 1

4.2. Filter Method

The method of extending the scale beyond 2,400° K. at the N.P.L. is to use the lamps calibrated against a furnace up to 2,400° K. together with a coloured glass filter. The filter can be yellow or blue in colour. If it is yellow (Harding 1944) it is placed in front of the lamp at the higher colour temperature, if it is blue it is placed in front of the lamp at the lower colour temperature. The filter has to be carefully made and its calibration is exacting because, for instance, an error of 1% in the spectral transmission for light of one wavelength may cause an error of 1° K. in the calculation of its colour temperature step.

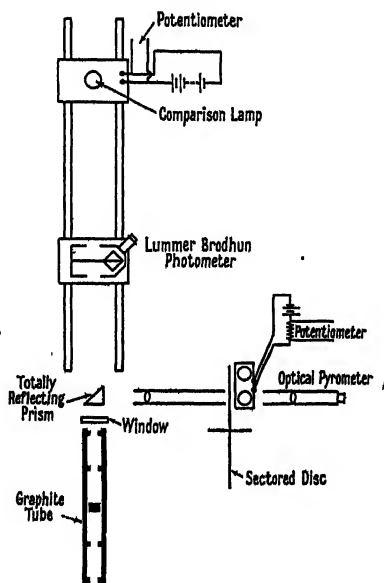


Figure 3.

Reproduced by kind permission of the International Commission on Illumination, New York.

The fact that a single filter can be used for the whole temperature scale can be illustrated by considering two lamps set up on a photometer bench with an ideal filter in front of one of them. The spectral distribution of the energy from the lamp with the filter in front of it is given by

$$E_{\lambda\theta_1} = \frac{t_\lambda C_1 \lambda^{-5}}{\exp(C_2/\lambda\theta_1) - 1}, \quad \dots\dots(3)$$

where t_λ is the spectral transmission of the filter, C_1 and C_2 are constants, λ is the wavelength of the light from the radiator and θ_1 is the lamp colour temperature.

The spectral distribution of energy from the bare lamp is

$$E_{\lambda\theta_2} = \frac{C_1 \lambda^{-5}}{\exp(C_2/\lambda\theta_2) - 1}. \quad \dots\dots(4)$$

Now for a colour match, energy match and intensity match at the photometer these spectral energies must be identical so that

$$t_\lambda \frac{C_1 \lambda^{-5}}{\exp(C_2/\lambda\theta_1) - 1} = \frac{C_1 \lambda^{-5}}{\exp(C_2/\lambda\theta_2) - 1}. \quad \dots\dots(5)$$

The shape of the spectral transmission curve is therefore given by

$$t_\lambda = \frac{\exp(C_2/\lambda\theta_1) - 1}{\exp(C_2/\lambda\theta_2) - 1} \simeq \exp\left\{\frac{C_2}{\lambda}\left(\frac{1}{\theta_1} - \frac{1}{\theta_2}\right)\right\}. \quad \dots\dots(6)$$

The omission of the -1 in Planck's formula means that Wien's formula is considered to be a good enough approximation to Planck's formula in this instance because for tungsten lamp temperatures $\exp(C_2/\lambda\theta)$ will always be greater than about 200. For actual filters the highest value for t_λ is arranged to be about 90%.

One filter then made with a spectral transmission curve to this specification will convert the spectral emission of a lamp at one temperature to that of a lamp at another temperature and it will do this whatever the temperature of one of the lamps, because temperature only occurs in the form $(1/\theta_1 - 1/\theta_2)$. The value of this, which is constant for each particular ideal filter, is called the reciprocal filter step. To save calculating reciprocals too frequently Priest (1933) introduced a new term, the Mired, which is a million divided by the colour temperature in degrees Kelvin, so that if a filter has a step of 100 mireds we know that

$$\frac{10^6}{\theta_1^\circ \text{K.}} - \frac{10^6}{\theta_2^\circ \text{K.}} = 100. \quad \dots\dots(7)$$

There is one other fact to be got from the formula for the spectral transmission. Taking logarithms for equation (6)

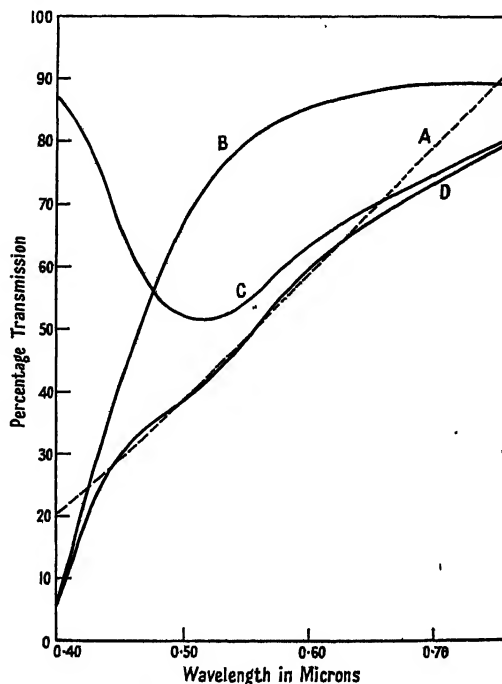
$$\log \frac{1}{t_\lambda} = \frac{C}{\lambda} \left(\frac{1}{\theta_2} - \frac{1}{\theta_1} \right) \log e. \quad \dots\dots(8)$$

As $\log(1/t_\lambda)$ is equal to the optical density of the filter the reciprocal filter step is proportional to the density, so that, for instance, two similar filters together would have double the reciprocal filter step of one alone. The reciprocal step of an ideal filter is approximately proportional to its thickness; this is not strictly true because reflection from the glass surfaces should be considered as well as the approximation of Wien's to Planck's formula.

The calculation of the spectral transmission of the filter is straightforward, but the realization of such a filter is difficult because it involves accurate spectrophotometry, accurate glass working and the reasonable selection of the coloured components. For work when one or two degrees accuracy is aimed at, at least two coloured components are required. One component generally has a spectral transmission curve about the required shape for the energy match and the other is generally a pale purple or green constituent to improve the colour match. Liquid filters are often of little value because of their lack of stability, diffusion at the container surfaces after a short time, and high temperature coefficients. Glass filters are generally made and even these have to be reconditioned by repolishing the surfaces after about five years because self-blooming alters the spectral transmission curves by a few tenths of one per cent and this can be sufficient to alter the colour by an amount that will introduce observer differences of several degrees. The properties of yellow filters (Harding 1944) made at the laboratory are illustrated by the next two figures.

Figure 4 shows the approximation of the actual to the ideal spectral transmission curve and Table 1 the colour-matching properties. With this filter the chromaticities of the filtered light are less than a fiftieth of the smallest perceptible difference from the Planckian chromaticities shown in brackets.

The equipment required to make the filter that has been considered is not generally available and it is unlikely that many potential users of such filters could get them. During the war there was a demand for filters of this kind and the Admiralty asked Messrs. Chance Brothers to produce large-scale melts of a blue



A Theoretical Filter. B Corning 346 2.490 mm. C Corning 507 0.370 mm. D Actual Filter.
Figure 4.

Table 1

Temperature of filter (° C.)	Illuminant (° K.)	Colour quality	Percentage transmission	Equivalent colour temperature (° K.)
20	3200	0.47664 X+0.41366 Y+0.10970 Z (0.47663 X+0.41364 Y+0.10973 Z)	52.2	2497.5
20	2848	0.49708 X+0.41526 Y+0.08766 Z (0.49708 X+0.41525 Y+0.08767 Z)	52.8	2279.7
20	2550	0.51680 X+0.41456 Y+0.06864 Z (0.51679 X+0.41454 Y+0.06867 Z)	53.4	2086.5
20	2360	0.53064 X+0.41269 Y+0.05667 Z (0.53063 X+0.41265 Y+0.05672 Z)	53.9	1958.8
31	2848	0.49729 X+0.41527 Y+0.08744 Z (0.49728 X+0.41526 Y+0.08746 Z)	52.6	2277.6

filter that would do for approximate colour temperature and intensity measurements. The N.P.L. indicated the required colorimetric properties and examined the various glass melts that were made. The result was Chance OB9 glass (Harding 1948). The original melt of glass was about twice as good for its purpose as any

other glass that we have examined and a subsequent melt has been even better. OB9 glass is not as good as the two-component filter made at the laboratory, but this could scarcely be expected as the improvements were discontinued as soon as the original colorimetric specifications for a filter for approximate measurements were satisfied.

The ratios of the spectral energies at a photometer when a lamp at $2,850^{\circ}\text{K.}$ (E_{2850}) is matched for chromaticity and luminance with a lamp at $2,250^{\circ}\text{K.}$ having an OB9 filter in front of it ($OB9_{2250}$) are shown in Table 2.

Table 2

Wavelength	$OB9_{2250}/E_{2850}$	Wavelength	$OB9_{2250}/E_{2850}$	Wavelength	$OB9_{2250}/E_{2850}$
0.40	0.89	0.54	0.97	0.68	1.16
0.42	0.97	0.56	1.04	0.70	1.31
0.44	1.01	0.58	1.04	0.72	1.32
0.46	1.01	0.60	1.01	0.74	1.30
0.48	0.98	0.62	0.97	0.76	1.29
0.50	0.97	0.64	0.90		
0.52	0.98	0.66	0.96		

Reproduced by kind permission of the Institute of Physics.

The residual greenness when colour matches are made with OB9 glass and tungsten filament lamps are indicated by the values of Δy in Table 3. The Δy value has been suggested by the author as a means of expressing the greenness of filters as compared with Planckian radiators. If x_1 and y_1 are chromaticity coordinates for a filter and x_1 and y_2 are those for a Planckian radiator that has the same coordinate x_1 as the filter, then $\Delta y = y_1 - y_2$.

Table 3. Values of Δy for given Thicknesses of OB9 Glass

Colour temperature of illuminant ($^{\circ}\text{K.}$)	Δy				
	Thickness of OB9				
	1 mm.	2 mm.	3 mm.	4 mm.	5 mm.
1500	+0.0008	+0.0018	+0.0028	+0.0036	+0.0038
2000	+0.0010	+0.0019	+0.0021	+0.0015	+0.0000
2500	+0.0008	+0.0012	+0.0006	-0.0006	-0.0021
3000	+0.0006	+0.0005	-0.0003	-0.0014	-0.0028
3500	+0.0004	+0.0001	-0.0007	-0.0018	

Reproduced by kind permission of the Institute of Physics.

(The reciprocal filter stop of OB9 glass is about 60 mireds/mm.)

A Δy value of 0.001 indicates a just perceptible chromaticity difference, as observed in a photometer with a ten degree Lummer-Brodhun field. It can be seen from this table that for the thicknesses of glass that are likely to be used for most colour-temperature measurements there will never be more than a just perceptible chromaticity difference.

Priest's method of measuring colour temperature might be mentioned here (Priest 1922) because it is essentially a filter method. A quartz plate of suitable thickness was placed between Nicol prisms, as illustrated in Figure 5. This arrangement produced a filter of a variable spectral transmission which is

characteristic of the ideal filter that has been considered previously, namely $t_\lambda = Ae^{B/\lambda}$, where A is an arbitrary constant and B is proportional to the difference of the reciprocal temperatures and to C_2 . Priest's equipment did not give this

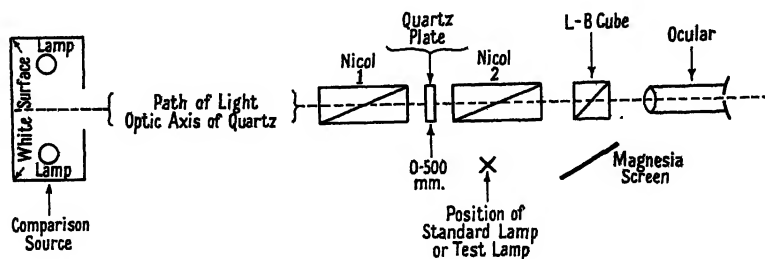


Figure 5.

After Priest, 1922. Reproduced by kind permission of the Optical Society of America.

precisely. By the use of this equipment Priest estimated that the colour temperature of tungsten at its melting point was 3,644° K. and that of a carbon arc with solid carbons was 3,780° K.

4.3. Selective Radiators

There are two reasons for considering the problem of selective radiators. The first is that it is often convenient to assign a colour temperature to light sources such as fluorescent lamps and artificial daylight fittings so that a single number is used to describe the chromaticity of the light. These light sources do not emit radiation with spectral energy distribution curves having the same shapes as those of full-radiator furnaces, and their chromaticities are very rarely identical with those of black bodies, so that some means of assigning colour temperatures has to be sought. A measurement of their chromaticities is a much more satisfactory procedure. The second reason is that when filters are made for colour-temperature measurements the chromaticity of the lamp and filter in combination is not always identical with that of a Planckian radiator. The colour-temperature steps of the filters have to be estimated for light sources at various colour temperatures and this cannot be done unless colour temperatures can be assigned to filter chromaticities which do not quite coincide with those of Planckian radiators.

Consider what is done when an attempt is made to use colour temperature to grade the chromaticities of illuminants. The chromaticities of the full radiators are plotted as well as the chromaticity of the light source for the standard C.I.E. observer. If the chromaticity happens to be on the full-radiator locus, the temperature of the full radiator that has this chromaticity can be read from the plot and so the equivalent colour temperature is obtained. The fact has, however, to be faced that if the estimate was put to a practical test and if the spectral emission of the light source was vastly different from that of the full radiator many real observers might not agree with the estimate to within the least measurable chromaticity difference.

If the point representing the chromaticity coordinates does not come on the full-radiator locus the position is worse. A point on the locus has to be chosen that will identify the chromaticity. This is done by applying the results of experiment to a calculation and the success of the assignment of a colour temperature is limited by observer differences.

Effects of small chromaticity differences.

The results of some of our experiments illustrate the effects of small chromaticity differences on colour temperature determinations. Our first experiment was to see what would happen if an attempt was made to colour-match a tungsten filament lamp with another of almost the same energy distribution, but with a slightly greener colour. The apparatus consisted of two amps on a photometer bench and an almost colourless glass filter resembling a piece of window glass. The lamps were first colour-matched at $2,400^{\circ}\text{K}$. then the piece of glass was placed in front of one of them and the voltage across the other was altered in an attempt to obtain a colour match between the unfiltered lamp and the filtered one. The results of this experiment are illustrated by Figure 6.

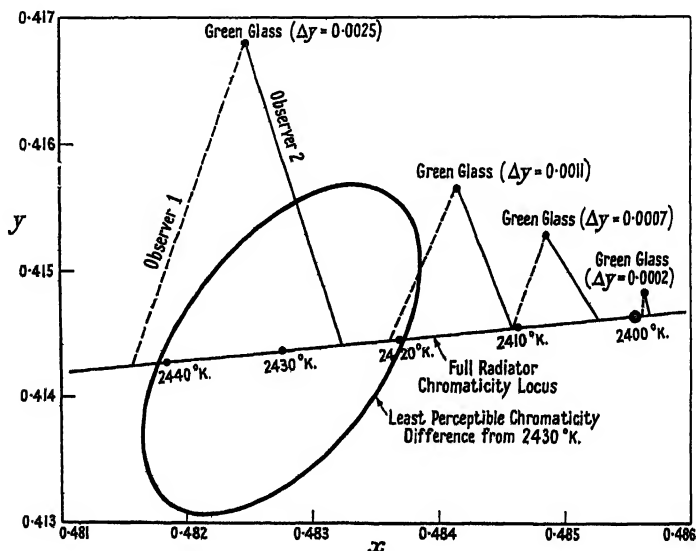


Figure 6. Observers' estimates of colour temperature for chromaticities not coincident with those of full radiators.

Observer 1 altered the colour temperature of the unfiltered lamp much more than observer 2 when the chromaticity difference was about the maximum tolerable. As the chromaticity differences were decreased by using different glass filters, the observers decided to estimate the best colour match in a certain way, which is indicated by drawing a line of defined gradient through the point representing the chromaticity to meet the full radiator chromaticity locus. These lines are often called iso-temperature lines (Judd 1936, MacAdam 1943). It is interesting to see that the observers maintain their method of estimation even when no chromaticity difference is perceptible. This experiment was repeated on different days and the observers repeated their previous settings to within about 2°K . and still maintained their 20°K . difference. The difference appears to be independent of the type of glass used to produce the greenness, that is the greenness may be produced by reducing the red or the blue light.

This observer difference has been put to the practical use of rejecting lamps that are slightly green. The two extreme observers of a group determine their observer differences for a filter of known greenness; they then compare lamps with bulbs known by previous measurement to be almost colourless with those to be inspected and if there is no observer difference the lamps are satisfactory,

if there is one, the observer difference gives a very approximate measure of the greenness.

The next experiment was to see what happens when there is a calculated colour match for the C.I.E. observer, but when the energy match is slightly different. Where the energy match is within about 5% over most of the visible spectrum and within about 20% at the red and blue ends observers were found to agree to better than 5° K.

The third experiment was to show how observers reacted to small chromaticity differences when an effort was made to measure the colour-temperature steps

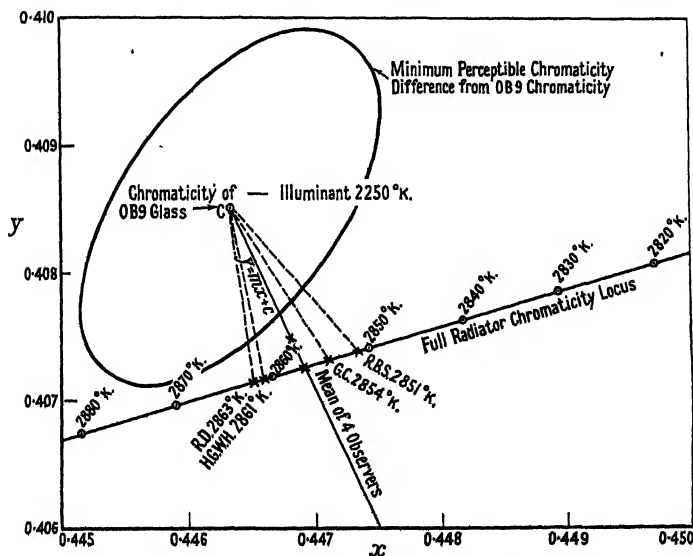


Figure 7. Gradient of the line to be drawn to estimate the colour temperature corresponding to the chromaticity of Chance OB9 glass.

of a slightly imperfect filter by means of lamps calibrated for colour temperature. The results that are quoted refer only to an OB9 blue glass filter with a colour-temperature step of about 2,250 to 2,850° K., but the procedure outlined can be used for any unknown filters. Two lamps and a Lummer-Brodhun photometer with a ten-degree field were set up on a bench. One lamp was operated at 2,250° K. and the blue filter was placed in front of it. The voltage across the terminals of the other lamp was varied until the best colour match was obtained. It was found that whilst each observer could repeat his settings to about 2° K. the observers differed by about 15° K. These results are illustrated in Figure 7.

4.4. Selective Absorption Methods

The methods of measuring colour temperature that have so far been considered have compared the whole spectral range of visible radiation from a light source with that from a standard lamp, and the measurements have therefore been made at the colour of some full radiator. If a light source had been slightly green this would have been noticed and steps taken to deal with the greenness, for instance the chromaticity may have been measured with a colorimeter because it was considered that a colour temperature could not be assigned to it. There are methods in which the whole of the spectral range of the visible light emitted by the light source is not considered which may be quite satisfactory for routine determinations, but care needs to be taken with them. They generally depend

on measuring the ratio of the intensity of the red light to that of the blue light from a source by letting the light fall on either a photoelectric cell or the eye, first when a red filter is placed in front of the source, then when a blue filter is placed there. Calibrated lamps are generally chosen as standards for determining the red to blue intensities for known colour temperatures. This method is satisfactory when the energy distributions are Planckian. If the energy distributions are not Planckian the results may have very little reference to the colour temperature, for instance, the method would give a red to blue ratio for a mercury or a sodium or a daylight lamp but the chromaticities corresponding to full radiators at these colour temperatures would be appreciably different from the actual chromaticities of the light sources. For light sources with a Planckian distribution this method can be made an absolute one if the spectral transmissions of the filters and the spectral sensitivity of the photoelectric cell are known, because the ratio of the red to blue responses can then be calculated for Planckian radiators at various temperatures. Photoelectric cell sensitivities are, however, difficult to measure with an accuracy that would ensure that the final result would be more accurate than a visual measurement made by a direct method in which the whole visible spectral range of light was used. Some errors can be avoided if the red to blue ratios are measured for one temperature within the range and the calculated scale then corrected according to the result, but a safer method is to use a comparison method with a standard lamp. The precision of these methods is generally about 10°K. in colour temperature for a one per cent change in red to blue response ratio, and different lamp bulbs may give results differing by as much as 10°K. at $2,800^{\circ}\text{K.}$ from the visual comparison.

The design of some photoelectric devices for red to blue ratio methods might be mentioned here. An attempt should be made to eliminate the effects of fluctuations of the intensity of the light source. This generally means that two cells have to be used, one covered with the red and the other covered with the blue filter. The cells are connected in opposition and a meter arranged to detect out-of-balance currents. The colour-temperature scale is provided by the positions of shutters arranged to vary the amount of light falling on the cells to obtain equality of output. If high precision is required, and especially if rectifier cells are used, it is desirable to arrange for the intensity of the incident radiation to be kept at a predetermined level. It is also desirable that all the incident light should pass through a filter having a high transmission for visible radiation, but falling to about 10% at 0.40μ and 0.70μ . This restricts the response of the cells to the visible spectrum range which is desirable for colour-temperature measurements and it also reduces the drifts that sometimes occur when red light falls on rectifier cells. The red and blue filters are chosen for their high transmissions for red or blue light. If the transmission band can be restricted to the red and blue regions of the spectrum only, the maximum sensitivity of the arrangement will be obtained. This cannot generally be achieved with actual filters, but fortunately the sensitivity is not greatly diminished if the filters do transmit some light in other spectral regions. If very dense red and blue filters are used and if the colour temperature of the illuminant is altered by a specified amount, the out-of-balance cell current will only be small because very little light will reach the cells. This arrangement would give the maximum red to blue response ratio, because the spectral bands transmitted could be those near to the red and blue ends of the spectrum. Alternatively, if the filters are

barely coloured, both clear glass for instance, each cell current will be large, because of the intensity of the incident light will be hardly reduced, but the out-of-balance current would be zero because the change of colour temperature of the illuminant will have the same effect on each cell if the spectral sensitivities are similar. There is therefore some value for the filter transmissions that will give maximum sensitivity. Calculations at the laboratory have shown, for example, that for filters of the types described used with rectifier cells the transmissions as measured with the cell and isolating filter should be between 25 and 50%.

A comparator made at the laboratory using selenium rectifier photoelectric cells with a galvanometer of sensitivity $150 \text{ mm}/\mu\text{a.}$ has a precision of about 2°K. for the tungsten filament lamp range. In this instrument the photoelectric cells are mounted side by side behind a hole in a sphere and shutters are arranged to cover the cells by varying amounts. One cell is covered with Chance OB3 and the other with Chance OY1 glass. OB3 glass is apparently no longer available, but Chance OB8 or OB2 glasses should provide a satisfactory alternative. Light enters the sphere through a hole in the wall opposite the cells and a disc at the centre of the sphere prevents direct light from falling on them. This entrance hole is covered with a filter made from Chance OY9 and Corning 978 glasses for isolating the visible spectrum, and an annular cell mounted just in front of the hole and behind the filter is used as an indicator for the intensity of the incident light. Chance OY9 is no longer available, but OY18 should provide a reasonable substitute.

A small portable two-cell colour-temperature meter with the same glass filters and a $15\text{--}0\text{--}15 \mu\text{a.}$ meter has a range from $1,800^\circ \text{K.}$ to infinite temperature marked on a scale 2 cm. long. The sensitivity is about 100°K. at $2,000^\circ \text{K.}$ and about $1,000^\circ \text{K.}$ at $10,000^\circ \text{K.}$ In order to calibrate this photoelectric meter for daylight colour temperatures a visual instrument has been made. This uses an OB9 glass wedge with auxiliary filters and the Δy values of the chromaticities of the filters are always less than 0.0025 for the colour-temperature range from $2,000^\circ \text{K.}$ to infinite temperature.

Red to blue ratio methods are particularly useful when the colour temperature of a light source, like a tungsten filament lamp, has to be determined when the lamp is in a sphere coated with a paint that is a good diffuser but is not strictly neutral in colour. If the light source is interchanged with a calibrated lamp the colour temperature of this standard lamp can be varied until the red to blue ratio is the same as that for the unknown light source and so the colour temperature of the unknown light source can be determined. The neutrality of the paint, the spectral sensitivity of the photoelectric cell and the spectral transmissions of the filters need not be known: all that has to be determined is the red to blue ratio for both light sources when they are in the sphere.

In another type of instrument which has been used for colour-temperature measurements some of the light from the source passes through a yellow filter and some through a dichroic filter which transmits red and green light. The ratio of the red to green intensities from the dichroic filter is altered by a colour wedge until the light from the yellow filter colour-matches that from the dichroic filter. Results obtained with this type of instrument depend on the observer's colour vision and an initial individual calibration is necessary if large errors are to be avoided.

4.5. *Indirect Methods*

One method, requiring expensive equipment, is to let the light from a lamp pass through a monochromator and be detected either visually or photo-electrically. If the lamp has been standardized for colour temperature when it is operated at a particular voltage the colour temperature at any other voltage can be obtained by changing the voltage to that required and determining the ratio of the energies emitted for radiation of two or three wavelengths, then finding by calculation from Planck's formula the temperature of the radiator that would give these ratios. Lamps can also be compared in this way, but it is advisable that the sensitivity of the equipment should be great enough for measuring the light after it has been allowed to fall on a diffusing medium such as magnesium oxide.

Another method of extending the colour temperature scale of a lamp, if its calibration is known for a few points, is to plot the logarithm of the power in watts against the logarithm of the colour temperature. The result is a straight line that can be extended beyond the known calibration. An experiment showed that if the colour temperature of a 500-watt or 1,000-watt lamp was determined for 2,000, 2,200 and 2,400° K. the 3,000° K. point could be estimated with an error of about 5° K. The values of the gradients of the lines joining the log (watts) and log (colour temperature) points for the 500-watt and 1,000-watt lamps were within 1% of each other.

§ 5. CONCLUSIONS

Colour temperature is very useful to express the operating conditions of light sources such as tungsten filaments, but should not be used indiscriminately for other light sources, because, if the colour and energy producing the colour are different from those of a full radiator, observer differences and minimum perceptible colour differences have to be considered before the value of colour temperature to specify the colour can be assessed.

ACKNOWLEDGMENT

Some of the work described above has been carried out as part of the general research programme of the National Physical Laboratory, and this paper is published by permission of the Director of the Laboratory.

REFERENCES

- BUCKLEY, H., COLLIER, L. J., and BROOKES, F. J. C., 1924, *Sixth Session of the Commission Internationale de L'Éclairage* (Cambridge: University Press, 1926), p. 203.
 C.I.P.M., 1948, *Échelle Internationale de Température de 1948, Procès Verbaux des Séances du Comité International des Poids et Mesures*, t. XXI, 1948.
 COBLENTZ, W. W., 1946, *Amer. Astr. Soc.*, 10, 222.
 FORSYTHE, W. E., 1923, *J. Opt. Soc. Amer.*, 7, 1115.
 HARDING, H. G. W., 1944, *Proc. Phys. Soc.*, 56, 21; 1948, *J. Sci. Instrum.*, 25, 333.
 HARDING, H. G. W., and SISSON, R. B., 1947, *Proc. Phys. Soc.*, 59, 814.
 JUDD, D. B., 1936, *J. Opt. Soc. Amer.*, 26, 421.
 MACADAM, D. L., 1943, *J. Opt. Soc. Amer.*, 33, 18.
 MAKOWSKI, M. W., 1949, *Rev. Sci. Instrum.*, 20, 876.
 PATERSON, C. C., and DUDDING, B. P., 1914-15, *Proc. Phys. Soc.*, 27, 230.
 PRIEST, I. G., 1922, *J. Opt. Soc. Amer.*, 6, 27; 1933, *Ibid.*, 23, 41.
 WORTHING, A. G., 1917, *Phys. Rev.*, 10, 377.

Certain Properties of Electrostatic Fields Encountered in Electron Lenses*

By P. A. LINDSAY
Imperial College, London†

MS. received 8th February 1950

ABSTRACT. The equipotential lines in a bipotential electron lens exhibit a fine structure which is revealed by application of the relaxation method to the solution of the Laplace equation in cylindrical coordinates. The asymmetry of the potential field between the two edges of the cylinders is confirmed by the same method and is established to be of the order of 2π in one particular case.

§1. INTRODUCTION

IT appears from some publications on electron optics (Cosslett 1946, Zworykin *et al.* 1945, Motz and Klanfer 1946) which have dealt with the potential distribution in a bipotential lens, that there are two details which, while probably familiar to many workers in electron optics, have not been stated explicitly. This is possibly due to considerable mathematical difficulties connected with the solution of electrostatic field equations for actual boundary conditions encountered in electron optics.

§2. FINE STRUCTURE OF THE EQUIPOTENTIAL LINES

During the use of the relaxation method (Southwell 1946, Allen *et al.* 1945) for theoretical investigations of the potential distribution in a bipotential lens consisting of two coaxial cylinders of equal diameter D and separated $D/8$ apart it was noticed that the equipotential curves change the sign of their curvature twice in the region between the axis of symmetry and the walls of the cylinders. Since no previous theoretical results indicated the existence of such apparent irregularity the results arrived at by the application of the relaxation method might have been questioned originally on the grounds of inaccuracies connected with the transition from the differential to the finite difference equations. However, this apparently unusual shape of the equipotential curves is clearly noticeable in the experimental results published by Maloff and Epstein (1938, p. 74). They do not mention this explicitly, possibly only due to the fact that the order of quantities involved just corresponds to the order of errors common in the method of electrolytic trough measurements. The electric potential function has been calculated in the present case to an accuracy within approximately 0.1%, which exceeds by almost two orders the accuracy of most previously published results (e.g. Bertram 1942). It was further established that increasing the number of points at which the value of the potential function was calculated and hence the overall accuracy of the results, indicated more and more clearly the existence of this double change in the sign of

* This paper forms part of a thesis submitted to the University of London for the degree of Ph.D.

† Now with Electric and Musical Industries Ltd., Hayes, Middx.

curvature of the equipotentials in the region under consideration. Figure 1 shows the equipotential lines as calculated by the relaxation method (full lines) compared with those obtained by applying Bertram's (1942) linear approximation.

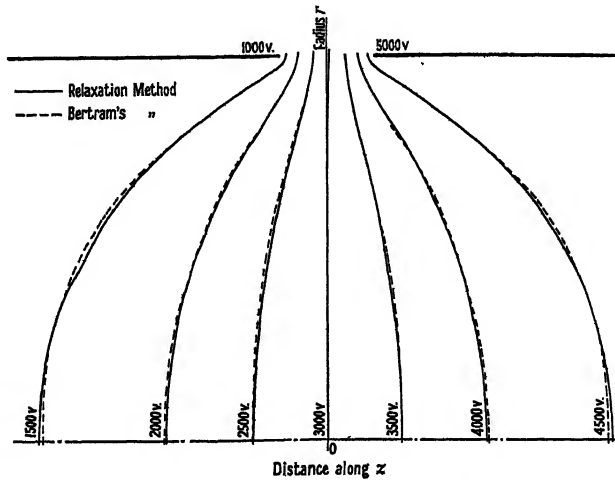


Figure 1. Equipotential lines.

It might be of interest to add that further investigations of the shape of the equipotentials in a bipotential lens indicate that the double change in the sign of curvature disappears in the limiting case of an infinitely small gap between the cylinders.

§ 3. GENERAL PROPERTIES OF THE POTENTIAL FUNCTION BETWEEN THE EDGES OF THE CYLINDERS

The other point of interest concerns the general properties of the potential function in a bipotential lens. It is seldom sufficiently emphasized that, taking the arithmetic mean of the cylinder potentials as a reference point, the potential function along the line connecting the inner edges of the two cylinders is not in general skew symmetrical. This can be proved easily by considering the additive properties of the potential function and constructing simple boundary conditions, which would add up to give actual potentials (e.g. 1,000 v. and 5,000 v.) applied across the cylinders.

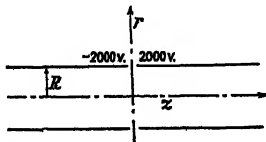


Figure 2. Boundary conditions in case I.

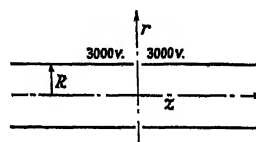


Figure 3. Boundary conditions in case II.

First consider two cylinders of equal but opposite potentials of $-2,000$ v. and $2,000$ v. as in Figure 2. Then for reasons of symmetry the equipotential line corresponding to the arithmetic mean of the cylinder potentials (in our case 0 v.) must coincide with the r axis, and the potential function between the two edges along the surface $r=R$, R being the cylinder radius, must be skew symmetrical relative to the plane $z=0$, as shown in Figure 4.

Next apply across both cylinders a potential of 3,000 v. as shown in Figure 3. This time the potential function along the surface $r = R$ must be symmetrical relative to the plane $z = 0$, because the conditions are identical on both sides of the plane $z = 0$. Yet the value of the potential function anywhere along $z = 0$ is less than 3,000 v. (at least as long as we assume that $\phi \rightarrow 0$ for $r, z \rightarrow \infty$). Adding cases I and II of Figure 4 in order to obtain 1,000 v. and 5,000 v. boundary conditions, it is obvious that the potential function between the edges of the two cylinders consists in general of the

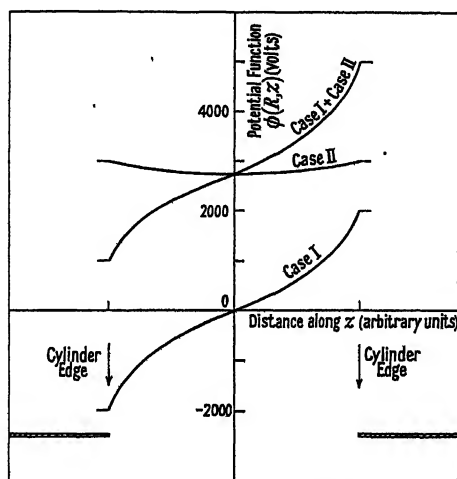


Figure 4. Potential distribution for two different boundary conditions.

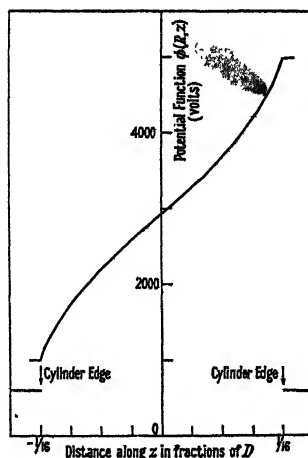


Figure 5. Potential distribution in the gap between cylinders.

z in fractions of D	$-\frac{1}{16}$	$-\frac{3}{64}$	$-\frac{1}{32}$	$-\frac{1}{64}$	0	$\frac{1}{64}$	$\frac{1}{32}$	$\frac{3}{64}$	$\frac{1}{16}$
$\phi(R, z)$ in volts	1000	1733	2199	2579	2932	3291	3588	4190	5000

sum of a symmetrical and a skew symmetrical function and hence its value would differ at $z = 0$ from the arithmetic mean of the cylinder potentials. Consequently the equipotential corresponding to this value of the potential will not be the plane of geometrical symmetry $z = 0$, but some other surface, slightly concave towards one of the cylinders, as shown in Figures 1 and 4.

Using ordinary analytical methods, it seems impossible to calculate the shape of the potential function in the space between the edges, because all the more drastic approximations usually introduced in the calculation of the potential distribution in a bipotential lens have to be applied of necessity to this particular region. Yet using the relaxation method and allowing for smaller accuracy around the edges of the cylinders it is still possible to show the general properties of the potential function in that part of the lens. The results support previous expectations and indicate clearly the non-symmetrical properties of the function. Making no claims to great accuracy of the results in this mathematically difficult region of the lens, it was shown in one particular case that the function at the point half way between the edges of the cylinders had a value that was short of the arithmetic mean by 2% (see Figure 5). This result indicates clearly that if we assume the function to be symmetrical in the gap between the edges of the cylinders, as is frequently done in extrapolating boundary conditions, as suggested by Cosslett (1946), Zworykin *et al.* (1945) and Bertram (1942), uncertainty of the resulting potential distribution calculations is at least of this order of magnitude.

REFERENCES

- ALLEN, D. N. DE G., SOUTHWELL, R. V., and VAISEY, G., 1945, *Proc. Roy. Soc. A*, **183**, 258.
BERTRAM, S., 1942, *J. Appl. Phys.*, **13**, 496.
COSSLETT, V. E., 1946, *Introduction to Electron Optics* (Oxford : Clarendon Press).
MALOFF, I. G., and EPSTEIN, D. W., 1938, *Electron Optics in Television* (New York : McGraw-Hill).
MOTZ, H., and KLANFER, L., 1946, *Proc. Phys. Soc.*, **58**, 30.
SOUTHWELL, R. V., 1946, *Relaxation Methods in Theoretical Physics* (Oxford : Clarendon Press).
ZWORYKIN, V. K., *et al.*, 1945, *Electron Optics and the Electron Microscope* (New York : John Wiley and Sons).

Reduction of the Spherical Aberration of Magnetic Electron Lenses

By U. F. GIANOLA

Department of Electron Physics, University of Birmingham

Communicated by J. Sayers; MS. received 18th January 1950

ABSTRACT. The dependence of the resolving power of the asymmetrical bell-shaped magnetic field on the lens parameters is examined for varying degrees of field asymmetry. It is shown that a high asymmetry index favours high resolution, and a preferable value of lens power is specified.

By suitably reinforcing the lens field with that of a small air-cored coil, the theoretical resolving power of the lens is increased. On account of the small size of the coil, the coil current must be pulsed intermittently if an appreciable reinforcing field is to be produced.

The basic factors governing the design and operation of a suitable coil are discussed.

§ 1. INTRODUCTION

THE spherical aberration inherent in high resolution electron lenses sets an ultimate limit to the resolving power of the electron microscope. Extensive researches, both theoretical and experimental (Scherzer 1947, Gabor 1949, and others), have not yet yielded practical means of correcting this aberration.

Nevertheless, it is theoretically possible to reduce the spherical error below any desired limit by a suitable choice of the form of the imaging field (Rebsch 1938). Unfortunately, the magnetic saturation of the pole-pieces sets a practical limitation to such improvements in the magnetic lens (Cosslett 1946, Liebmann 1946, Glaser 1949), although Cosslett has shown that an appreciable gain can be obtained by an appropriate increase in accelerating voltage and lens dimensions.

The possible diminution of the aperture defect by the introduction of localized fields, produced by subsidiary coils, into existing lenses will be considered in the following for certain idealized instances.

§ 2. THE LIMIT OF RESOLUTION OF MAGNETIC LENSES

The axial fields typical of practical magnetic lenses may be closely approximated by the asymmetrical bell-shaped fields which have been the subject of a rigorous analytical treatment by Glaser (1940) and Dosse (1940), who gave, in particular, the spherical and chromatic aberration constants, C_s and C_c respectively, in terms of the lens power parameters k_a and k_b :

$$k_a^2 = ea^2H_0^2/8mU; \quad k_b^2 = eb^2H_0^2/8mU. \quad \dots\dots(1)$$

Here $2a$ and $2b$ are the widths of the two component symmetrical distributions at half-field strength, H_0 is the central field strength, and U is the accelerating voltage.

Dosse (1940) has evaluated these aberration constants for the object position giving high magnification, for field asymmetries of degree $q = a/b = k_a/k_b$.

In the case of the Glaser distribution, the minimum error (Glaser 1949) may be expressed in terms of the independent lens parameters :

$$d_0 \simeq [C_s(U, b, H_0)\lambda^3(U)]^{1/4} \quad \text{for } \alpha = \alpha_0 \simeq \lambda/d_0; \quad \dots\dots(2)$$

λ is the de Broglie wavelength of the illuminating electrons.

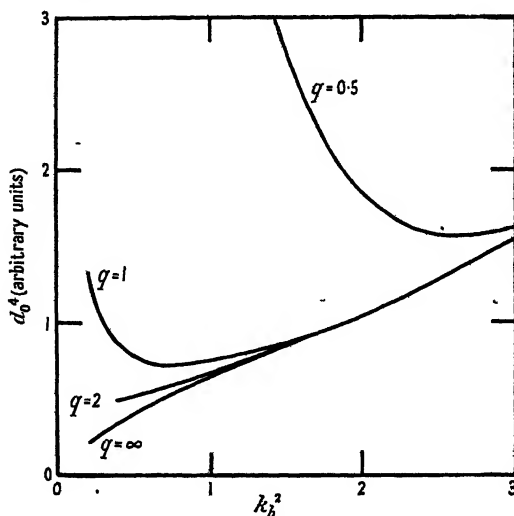


Figure 1. Variation of the minimum error with lens power at constant field strength and half-field width.

$$H_0 \text{ and } b \text{ constant; } k_b^2 \propto 1/U; \quad d_0^4 \propto C_s k_b^3.$$

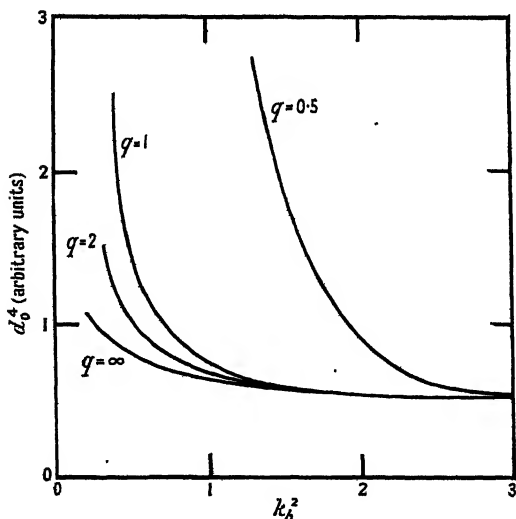


Figure 2. Variation of the minimum error with lens power at constant accelerating voltage and field strength.

$$U \text{ and } H_0 \text{ constant; } k_b^2 \propto b^2; \quad d_0^4 \propto \frac{C_s}{b} k_b.$$

The specific dependence of this error function on each of these parameters individually, under high magnification conditions, is illustrated in Figures 1, 2 and 3.

It will be noted that, quite generally, the minimum error is smaller for field distributions of large asymmetry factor q , although the gain over that of the symmetrical field (i.e. $q=1$) is small.

For distributions of small asymmetry factor, that is of the order of unity and less, an optimum value of voltage exists for predetermined values of b and H_0 (Figure 1), as has been shown previously (Cosslett 1946, Liebmann 1946). However, for large asymmetry factors no such optimum appears for practical values of lens power, the error decreasing progressively with accelerating voltage.

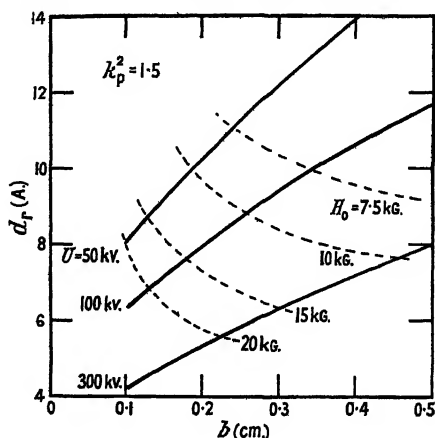


Figure 3. Variation of the ultimate limit of the resolving power of the bell-shaped field electron lens with half-field width at specific accelerating voltages. The power of the lens is kept at the preferred value $k_p^2=1.5$; the loci of specific field strength values are also shown. The chromatic error is calculated here for a voltage ripple of 2.5×10^{-5} , and an initial energy of thermionic emission of 1 ev.

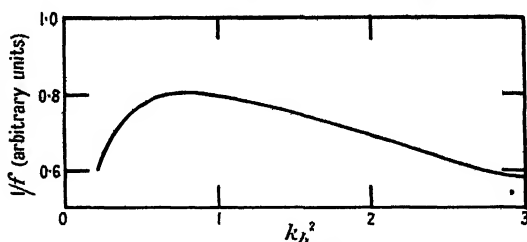


Figure 4. Variation of refractive power with the field distribution half-width only. U and H_0 constant.

A further increase in resolution is obtained when the half-field width only is increased (Figure 2). Since it is desirable to keep the dimensions of the electron-optical system within practical limits, it is preferable to adjust the lens power to keep the focal length of the objective small. As is illustrated in Figure 4, a minimum focal length is obtained at an optimum value of the half-field width. However, the fractional variation over a range of lens power is so small, for example $(f)_{k^2=2} = 1.17(f)_{k^2=0.8}$, so that this factor may be neglected. Any increase in lens power above about $k^2=1.5$ does not result in an appreciable diminution of the minimum error. Hence, in the following the value $k_p^2=1.5$ will be adopted as a preferential lens power. In addition, for this preferential value of lens power, the small differences in the characteristics of different asymmetric distributions may be neglected.

A more rapid decrease in the error function with change in lens power results when the field strength only is varied, as is illustrated in Figure 3, which shows the decrease in the total error d_r (Zworykin *et al.* 1945) consequent on a decrease in the half-field width, accompanied by a corresponding increase in the field strength to maintain the lens power at the preferred value k_p^2 . The loci of the appropriate field strength parameter are given for specific values.

It follows from these considerations that the resolving power of existing lenses may be further improved at a determined voltage, by an appropriate increase in field strength and a reduction in lens dimensions. Unfortunately, as already indicated (Cosslett 1947), this course is restricted by magnetic saturation of the pole-pieces. Nevertheless, some progress in this direction may still be made by the introduction of subsidiary coils into existing lenses, or by the use of an air-cored coil alone as an electron lens.

§3. THE MAGNETIC FIELD PRODUCED BY AN AIR-CORED COIL

The magnetic field produced by an air-cored coil of axial symmetry may be expressed :

$$H(0, z) = 2\pi i \int_{-\infty}^{\infty} dz_0 \int_{r_1(z_0)}^{r_2(z_0)} \frac{r^2}{[r^2 + (z - z_0)^2]^{3/2}} dr, \quad \dots\dots(3)$$

where i is the current through unit cross section of the coil, assumed uniform, and r_1 and r_2 are the radial boundaries of the coil at a particular axial point z_0 .

For simplicity only coils of rectangular section (Figure 5) will be considered

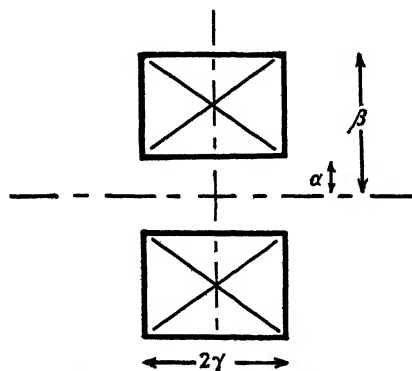


Figure 5. Cross-sectional diagram showing the dimensions of the air-cored coil considered.

here. The axial field strength is then given by :

$$\left. \begin{aligned} H(0, z) &= 2\pi i \alpha [k_1 \eta(h, k_1) - k_2 \eta(h, k_2)] \\ \eta(h, k) &= \ln \left[\frac{h + (h^2 + k^2)^{1/2}}{1 + (1 + k^2)^{1/2}} \right] \end{aligned} \right\} \quad \dots\dots(4)$$

where $h = \beta/\alpha$; $k_1 = (z + \gamma)/\alpha$; $k_2 = (z - \gamma)/\alpha$.

This field may be treated in accordance with the Glaser distribution for a suitable choice of coil dimensions.

In the following it will be assumed that the axial fields H_1 and H_2 , produced by the main lens and a subsidiary coil, may be represented by the symmetrical

Glaser distributions of half-widths $2a_1$ and $2a_2$, and that the respective radial axes are coincident. The resultant axial field may then be written

$$h_z = \frac{h_0}{1+s} \left[\frac{1+s+u(r+s)}{(1+u)(1+ru)} \right], \quad \dots (5)$$

where $r = (a_1/a_2)^2$; $s = H_2(0, 0)/H_1(0, 0)$; $u = (z/a_1)^2$.

It is assumed here that the influence of the subsidiary field on the magnetic saturation of the main field is small. This is justified if the relative dimensions of the coil are small.

An examination of equation (5) shows that the resultant axial field distribution will be given by that of equation (1) only if $a_1/a_2 = 1$, in which case $(h_z)_{r=1} = (1+s)H_1$. However, the resultant field may in general be considered in terms of the Glaser distribution to a reasonable approximation. It follows that, by reinforcing the main lens with a suitable subsidiary field, or by using an air-cored coil alone, a resultant field of increased magnitude and small half-width may be effected, and, in accordance with the above considerations, an improvement in resolution at a given voltage should be obtained.

In the absence of analyses of the electron-optical properties of other field distributions, the results of other possible combinations of main and subsidiary fields will need to be dealt with by numerical integration of the ray equation, or by direct experiment.

§ 4. PRACTICAL CONSIDERATIONS

The practical dimensions of air-cored coils needed to produce the compact fields characteristic of high power electron lenses are necessarily small. The average coil radius should be of the order of the required field half-width b , with cross-sectional area of the order of b^2 . Thus the construction of a suitable coil without introducing lens asymmetries would not prove easy. Fortunately, such asymmetries may be compensated by the use of correcting cylindrical elements (Scherzer 1947, Hillier and Ramberg 1947).

In addition, in view of the small cross section of the coil, it would not be possible to use steady currents to produce the several thousand gauss necessary to reduce the lens aberrations by an appreciable factor. However, by intermittently pulsing the coil current, magnetic fields of any reasonable magnitude may be obtained, and may indeed exceed the main lens field strength by a considerable factor. In practice, fields of the required form up to 100,000 gauss have been obtained without appreciable difficulty.

Assuming that adiabatic conditions prevail during the pulse, for an allowable temperature rise ΔT in the coil the maximum pulse length τ may be written

$$\tau = \frac{J\rho\sigma}{\mu} \frac{\Delta T}{i^2}, \quad \dots (6)$$

where ρ , σ , and μ are the density, specific heat and resistivity of the coil winding, J is the electrical equivalent of heat, and i is the current density in e.m.u./cm².

An allowable temperature rise of 150° c. in an enamel-coated copper wire coil gives the approximate relation

$$\tau \sim \frac{2}{\pi} \times 10^6 \text{ second,}$$

with a maximum repetition frequency ν

$$\nu \sim \frac{3S}{(\beta^2 - \alpha^2)\gamma^2} \times 10^{-3},$$

where S denotes the total coil surface area active in dissipating heat.

As a practical example, a coil of dimensions, $2\alpha = 0.2$ cm., $2\beta = 0.32$ cm., $2\gamma = 0.12$ cm. will produce a field with a half-width b of 1 mm. This coil may be made of 180 turns of 46 s.w.g. enamelled copper wire; then a 20 amp. current pulse will produce an axial field maximum of 20,000 gauss, and the maximum pulse length is 500 μ sec. with a repetition frequency of 5–10 pulses per second.*

Pulsed operation of electron lenses would severely increase the complexity of a microscope system, and, more important, would result in a reduced average image intensity and in correspondingly longer exposure times, with the existing techniques of image observation and recording. Consequently, the probability of image displacement, due to vibration and thermal drift, would be increased. However, an appreciable improvement in the efficiency of electron-image conversion should follow on using long afterglow luminescent screens coupled with an image amplifier. Pulsed operation of the electron source will also result in an intensity increase.

Other difficulties arise from the transient thermal expansion of the coil and from transient astigmatisms. Thus, in view of the many complications, the introduction of pulsed lenses may not prove feasible. However, considerable complications also accompany other methods suggested for the reduction of the aberrations of electron lenses (Scherzer 1947, Zworykin *et al.* 1945). Thus, in spite of the inherent practical difficulties, an air-cored lens on the lines indicated may serve a useful function in the development of electron microscopes of higher resolving power.

ACKNOWLEDGMENTS

The author is indebted to Professor J. Sayers for his encouragement throughout this work, which was made possible by a grant from the Department of Scientific and Industrial Research.

REFERENCES

- COSSLETT, V. E., 1946, *Proc. Phys. Soc.*, **58**, 443; 1947, *J. Sci. Instrum.*, **24**, 40.
 DOSSE, J., 1940, *Z. Phys.*, **117**, 316.
 GABOR, D., 1949, *The Electron Microscope*, 2nd edition (London: Electronic Engineering).
 GLASER, W., 1940, *Z. Phys.*, **117**, 285; 1949, *Acta Phys. Austriaca*, **3**, 38.
 HILLIER, J., and RAMBERG, E. G., 1947, *J. Appl. Phys.*, **18**, 48.
 LIEBMANN, G., 1946, *Phil. Mag.*, **37**, 677.
 SCHERZER, O., 1947, *Optik*, **2**, 114.
 REBSCH, R., 1938, *Ann. Phys., Lpz.*, **31**, 551.
 ZWORYKIN, V. K., *et al.*, 1945, *Electron Optics and the Electron Microscope* (New York: John Wiley).

* Investigation has shown that stabilization of the current pulse to 1 in 10^4 should not be difficult.

The Computation of Wave-Front Aberrations of Oblique Pencils in a Symmetrical Optical System

By W. WEINSTEIN

Technical Optics Section, Imperial College, London

MS. received 19th October 1949 and in amended form 23rd January 1950

ABSTRACT. Aberration formulae given previously, together with new ones for astigmatism and secondary spherical aberration, are put into forms which are suitable for computation and simplify the treatment of field lenses. A method of comparing the formulae with ray tracing is given, and the results for a triplet objective are discussed.

§ 1. INTRODUCTION

IN a previous paper (Weinstein 1949, referred to here as I) the author gave equations for the aberration coefficients of a pencil round a principal ray at a finite field angle, the coefficients being those for aberrations which vary as the fourth and third powers of the aperture. To these can be added equations for the secondary axial spherical aberration coefficient (see § 2 below), this being an approximation to the sixth power spherical aberration coefficient for oblique pencils. The total of results obtained, namely sagittal and meridian focal distances, third and fourth power oblique aberration coefficients and sixth power axial aberration coefficient, gives an approximate indication of the aberrations of an oblique pencil.

In order to test this, the aberration coefficients of a number of optical designs were calculated and the wave-front aberrations predicted by the coefficients were compared with those calculated from ray-tracing results, these being obtained by numerical integration of the transverse aberrations of various skew and meridian rays of the pencil.

For purposes of computation convenient formulae were derived from the fundamental aberration equations given in I and in § 2 below; also a method of computing astigmatism along a principal ray was used which involves tracing only an s -fan accurately, the quantity $1/t - 1/s$, which is of aberrational magnitude, being then computed to, say, three significant figures.

These computation formulae and some of the numerical results are given in this paper.

§ 2. AXIAL SECONDARY SPHERICAL ABERRATION

In this section refraction and transfer formulae for axial secondary spherical aberration are given.

Take rectangular axes with the origin O at the pole of a refracting surface, Ox along the axis of the system, and Oz perpendicular to the axis. Then the equation of the section by the (x, z) plane of a wave front having primary and secondary spherical aberration may be written

$$x = \frac{1}{2l} z^2 + \left(\frac{1}{8l^3} + C_4 \right) z^4 + \left(\frac{1}{16l^5} + C_6 \right) z^6 + O(z^8), \quad \dots\dots(1)$$

where l is the radius of curvature of the wave front at the point O and C_4 and C_6 are coefficients of primary and secondary spherical aberration. When C_4 and C_6 vanish equation (1) reduces to that of a sphere of radius l , neglecting quantities which are of the order of z^8 , so that the optical path difference between this sphere and the wave front is $N\{C_4 z^4 + C_6 z^6 + O(z^8)\}$, where N is the refractive index of the medium; this expression is therefore the spherical aberration at semi-aperture z , referred to the paraxial focus.

The refraction and transfer formulae for C_4 and C_6 can be derived by the methods used in I.

Refraction and transfer formulae for C_4 :

$$\Delta(NC_4) = \frac{1}{8}\Delta\left\{\frac{N}{l}\left(\frac{1}{r} - \frac{1}{l}\right)^2\right\}, \quad \dots\dots(2)$$

$$C_4^{(+1)} = \left(\frac{l'}{l(+1)}\right)^4 C_4'. \quad \dots\dots(3)$$

Refraction and transfer formulae for C_6 :

$$\Delta(NC_6) = \Delta\left\{N\left\{\frac{5}{2}\left(\frac{1}{r} - \frac{1}{l}\right)^2 - \frac{3}{r}\left(\frac{1}{r} - \frac{1}{l}\right) + \frac{1}{r^2}\right\}C_4 + \frac{N}{16l^2}\left(\frac{1}{r} - \frac{1}{l}\right)^3\right\}, \quad \dots\dots(4)$$

$$C_6^{(+1)} = \left(\frac{l'}{l(+1)}\right)^6 \left\{C_6' + \frac{8l'd'}{l' - d'} C_4'^2\right\}. \quad \dots\dots(5)$$

In these formulae Δ denotes the change in a quantity on refraction, primes denote quantities after refraction and the superscript $(+1)$ denotes the quantity just before refraction at the next surface; r is the radius of curvature of the refracting surface and N and N' the refractive indices of the two media.

The derivations of equations (2) to (5) will not be given here since (2) and (3) are equivalent to the ordinary primary aberration formulae, while a formula equivalent to (4) and (5), but using different variables, has been given by von Rohr (1920, p. 240).

§ 3. ASTIGMATISM ALONG A PRINCIPAL RAY

Young's formulae (see, for example, Conrady 1929, p. 407) for computing astigmatism when the principal ray has finite angles of incidence are

$$\Delta\left\{N\left(\frac{\cos^2 I}{t} - \frac{\cos I}{r}\right)\right\} = 0, \quad \dots\dots(6)$$

$$\Delta\left\{N\left(\frac{1}{s} - \frac{\cos I}{r}\right)\right\} = 0, \quad \dots\dots(7)$$

where s and t are the distances along the principal ray to the sagittal and meridian foci and I is the angle of incidence. Each of these formulae must be computed with the full precision of an ordinary paraxial ray-trace, say five or six significant figures, in order to obtain the astigmatic difference at the end of the calculation with sufficient accuracy. The following method removes the necessity for this precision with equation (7).

The equation of a wave front given in equation (2) of paper I, can be written in the form

$$x = \frac{1}{2s}(y^2 + z^2) + C_{20}y^2 + O(p^3), \quad \dots\dots(8)$$

where p is a measure of the aperture and

$$C_{20} = \frac{1}{2} \left(\frac{1}{t} - \frac{1}{s} \right), \quad \dots\dots(9)$$

so that C_{20} is the astigmatism coefficient.

From (6) and (7)

$$\Delta\{NC_{20} \cos^2 I\} = \frac{1}{2} N^2 \sin^2 I \Delta \left(\frac{1}{Ns} \right); \quad \dots\dots(10)$$

also, since $s^{(+1)} = s' - D'$, $t^{(+1)} = t' - D'$,

$$C_{20}^{(+1)} = \left(\frac{s'}{s^{(+1)}} \right)^2 \frac{1}{1 - \frac{2D's'}{s^{(+1)}} C_{20}'} C_{20}'. \quad \dots\dots(11)$$

Equations (10) and (11) are the refraction and transfer formulae for C_{20} , and they can be applied when s and s' have been obtained by the use of equation (6), with the precision usual in aberration computations, e.g. three or four significant figures.

§ 4. COMPUTING EQUATIONS

The refraction and transfer formulae tabulated in § 8 of paper I and equations (2) to (5) of this paper give directly the values of the aberration coefficients of a pencil in terms of the constructional data of the system and the lower order properties of the pencil; for example s - and t -intersection lengths and coma coefficients are involved in the fourth order equations. Since the coefficients uniquely characterize the structure of the pencil in the neighbourhood of the principal ray, these equations may be considered to be in their most fundamental forms. However, for computing purposes it is more convenient to use equations involving actual wave-front aberrations instead of aberration coefficients, and these equations will be derived in the sub-sections which follow.

(i) *Axial Aberrations*

In deriving the computing equations for the axial aberrations the notation of Conrady (1929 p.38) for paraxial ray-tracing will be used, with the additional symbol $\alpha = u + i$ for the paraxial 'central angle'. If y is the incidence height, at the surface considered, of the paraxial ray from the axial object point, the primary and secondary wave-front aberrations to be computed are defined by

$$W_4 = NC_4 y^4 \quad \dots\dots(12)$$

$$W_6 = NC_6 y^6 \quad \dots\dots(13)$$

respectively. Then if equation (2) is multiplied by y^4 and equation (3) by N the following computing formulae for W_4 are obtained:

$$\Delta W_4 = \frac{1}{2} y (Nz)^2 \Delta \left(\frac{u}{N} \right), \quad \dots\dots(14)$$

$$\Delta_{tr} W_4 = 0, \quad \dots\dots(15)$$

where Δ_{tr} denotes the increment on transfer between surfaces.

Similarly, on multiplying equation (4) by y^6 and equation (5) by N the computing formulae for W_6 are seen to be

$$\Delta W_6 = \frac{5}{2}(Ni)^2 \Delta \left(\frac{W_4}{N^2} \right) - 3\alpha Ni \Delta \left(\frac{W_4}{N} \right) + \alpha^2 \Delta W_4 + \frac{1}{16} y (Ni)^3 \Delta \left(\frac{u^2}{N^2} \right), \quad \dots (16)$$

$$\Delta_{tr} W_6 = \frac{8d'}{N'y'y^{(+1)}} W_4'^2. \quad \dots (17)$$

By means of these computing equations the successive increments of W_4 and W_6 on refraction and transfer at each surface can be added, to obtain the values at the last surface in the system; finally the value of W_6 at the exit pupil can be obtained by applying equation (17), and then the values of C_4 and C_6 at the pupil can be obtained from equations (12) and (13).

The significance of W_6 can be explained as follows:

Let S_0 denote the paraxial ray of which the data are used in the above computations and let S denote the 'marginal' ray which coincided with S_0 at the first surface. The values of W_6 found for each surface are the secondary wave-front aberrations at points on the wave front of which the distances from the axis are the incidence heights of S_0 ; these are not, however, equal to the incidence heights of S after the first surface so that these values of W_6 do not refer to the points in which S meets the wave fronts. It can easily be shown that the difference, although usually numerically small, is mathematically of the order of magnitude of W_6 , so that the values of W_6 do not refer to the ray S ; they are, in fact, the aberrations at the incidence heights of S_0 . Thus W_6 is primarily an auxiliary quantity which is used in computing the values of C_6 , the aberration coefficient.

(ii) Astigmatism

In defining the wave-front aberrations which are used as auxiliary computing functions for the oblique aberration coefficients, certain convergence angles, etc. analogous to the paraxial u , i , α , y are used. These variables were defined by Hopkins (1946), who used them to give computing formulae, with checks, for tracing s - and t -fans along a principal ray. In this paper Hopkins' notation will be used with the addition of subscripts s or t to denote quantities referring to the sagittal or tangential sections respectively.

Formulae giving Hopkins' variables in terms of the results of an astigmatism calculation by means of Young's formulae can easily be obtained, but it is preferable either to calculate both the s - and t -traces directly by Hopkins' method or to calculate the s -trace only by Hopkins' method and then apply equations (10) and (11) to determine the astigmatism coefficient C_{20} ; Hopkins' variables for the t -section can then be obtained by the following formulae:

$$\left. \begin{aligned} y_t' &= y_t \frac{\cos I'}{\cos I}, \\ u_t' &= \frac{y_t'}{s'} + 2y_t' C_{20}', \\ y_t^{(+1)} &= y_t' - D' u_t'. \end{aligned} \right\} \quad \dots (18)$$

The latter method has the advantage that part of the calculation (equations (10) and (11)) is carried out with quantities of aberrational magnitude.

(iii) *Third Degree Oblique Aberrations*

It is convenient to define the following symbols for use in this section and the next, all denoting quantities invariant on refraction :

$$\left. \begin{aligned} A &= N i_t \cos I, & A_s &= N i_s, \\ h &= \frac{y_t}{\cos I}, & G &= \frac{y_s}{h}, \\ B &= N \sin I. \end{aligned} \right\} \dots\dots(19)$$

A quantity which refers to an aberration of the n th degree, i.e. varying as the n th power of the aperture, will have a subscript ij , where $i+j=n$; also a bracketed superscript (m) will denote a quantity referring to the m th refracting surface. Then the third degree aberrations which are used as computing functions are W_{30} and W_{12} , defined by

$$W_{30} = N C_{30} y_t^3, \dots\dots(20)$$

$$W_{12} = N C_{12} y_t y_s^2, \dots\dots(21)$$

where C_{30} and C_{12} are aberration coefficients as defined in I. If equations (13) and (28) of paper I are multiplied by $(y_t/\cos I)^3$ and $y_t y_s^2/\cos I$ respectively they become

$$\Delta W_{30} = \frac{1}{2} B h A \Delta \left(\frac{u_t}{N \cos I} \right), \dots\dots(22)$$

$$\Delta W_{12} = \frac{1}{2} B \frac{y_s^2}{r} \Delta(u_t) - \frac{1}{2} B h \Delta(u_s^2). \dots\dots(23)$$

By equations (36) and (37) of paper I $\Delta_{tr} W_{30} = 0$ and $\Delta_{tr} W_{12} = 0$, so that equations (22) and (23) are the computing formulae for W_{30} and W_{12} ; if the system contains, say, k surfaces, the increments for each surface obtained from these equations are added to give $W_{30}^{(k)}$ and $W_{12}^{(k)}$; $C_{30}^{(k)}$ and $C_{12}^{(k)}$ can be obtained from these by means of equations (20) and (21), when they will refer to the wave front at the point at which the principal ray emerges from the last surface. Alternatively, the values of C_{30} and C_{12} can be made to refer to the wave front at the point where the principal ray in the image space crosses the axis, by using the transfer formulae of I.

(iv) *Fourth Degree Oblique Aberrations*

The wave-front aberrations used for the computing equations are defined by

$$W_{40} = N C_{40} y_t^4, \dots\dots(24)$$

$$W_{22} = N C_{22} y_t^2 y_s^2, \dots\dots(25)$$

$$W_{04} = N C_{04} y_s^4. \dots\dots(26)$$

The procedure for deducing the equations is similar to that outlined in (iii), and the results are as follows:

Refraction equation for W_{40} :

$$\begin{aligned} \Delta W_{40} &= \frac{1}{2} h A^2 \Delta \left(\frac{u_t}{N \cos I} \right) + B \left[3 A \Delta \left(\frac{W_{30}}{N^2 \cos^2 I} \right) - \frac{3}{2} \alpha_t \Delta \left(\frac{W_{30}}{N \cos I} \right) \right] \\ &+ B^2 \left[\frac{1}{2} h A \Delta \left(\frac{u_t^2}{N^2 \cos^2 I} \right) - \frac{1}{2} h \alpha_t^2 \Delta \left(\frac{u_t}{N \cos I} \right) \right]. \dots\dots(27) \end{aligned}$$

Transfer equation for W_{40} :

$$\Delta_{tr} W_{40} = \frac{9}{2} \frac{D'}{N' y_t' y_t^{(+1)}} W_{30}'^2. \quad \dots\dots(28)$$

Refraction equation for W_{22} :

$$\begin{aligned} \Delta W_{22} = & \frac{1}{2} A \frac{y_s^2}{r} \Delta(u_t) - \frac{1}{2} h A \Delta(u_s^2) + B \left[\frac{3}{2} \frac{y_s G}{r} \Delta\left(\frac{W_{30}}{N \cos I}\right) + A \Delta\left(\frac{W_{12}}{N^2 \cos^2 I}\right) \right. \\ & \left. - \frac{1}{2} \alpha_t \Delta\left(\frac{W_{12}}{N \cos I}\right) - 2 \alpha_t \Delta\left(\frac{W_{12} \cos I}{N}\right) + 2 \frac{A_s}{G} \Delta\left(\frac{W_{12}}{N^2}\right) \right] \\ & + B^2 \left[\frac{1}{4} \frac{y_s^2 \alpha_t}{r} \Delta\left(\frac{u_t}{N \cos I}\right) - \frac{1}{2} A \frac{y_s^2}{r} \Delta\left(\frac{u_t}{N^2 \cos^2 I}\right) - \frac{1}{2} \frac{h}{G} \Delta\left(\frac{u_s^3}{N}\right) \right]. \end{aligned} \quad \dots\dots(29)$$

Transfer equation for W_{22} :

$$\Delta_{tr} W_{22} = \frac{3 D'}{N' y_t' y_t^{(+1)}} W_{30}' W_{12}' + \frac{2 D'}{N' y_s' y_s^{(+1)}} W_{12}'^2. \quad \dots\dots(30)$$

Refraction equation for W_{04} :

$$\Delta W_{04} = \frac{1}{8} y_s A_s^2 \Delta\left(\frac{u_s}{N}\right) + \frac{1}{2} \frac{B y_s G}{r} \Delta\left(\frac{W_{12}}{N \cos I}\right) + \frac{1}{8} B^2 \frac{y_s^3}{r^2} \left[\Delta\left(\frac{u_s}{N}\right) - G \Delta\left(\frac{u_t}{N \cos I}\right) \right]. \quad \dots\dots(31)$$

Transfer equation for W_{04} :

$$\Delta_{tr} W_{04} = \frac{1}{2} \frac{D'}{N' y_t' y_t^{(+1)}} W_{12}'^2. \quad \dots\dots(32)$$

The successive increments for refraction and transfer at each surface are added to give the aberrations after refraction at the last surface; if desired the transfer equations can be used to find the values of W_{40} , W_{22} and W_{04} at the point where the principal ray crosses the axis in the image space; the values of the aberration coefficients for this point can then be found from equations (24) to (26), using appropriate values of y_t and y_s .

It should again be noted that, as in the case of W_6 , the quantities $W_{40} \dots W_{04}$ are not the aberrations of their respective types for one ray of the pencil throughout the system; they are the aberrations at points on the wave fronts of which the rectangular coordinates are (x, y_t, y_s) in the coordinate system used in I, and the differences between the aberrations at these points and those at the intersection points of an actual ray are not negligible. Thus W_{40} , W_{22} and W_{04} are auxiliary quantities, used for convenience in computing the aberration coefficients.

(v) Field Lenses

In field lenses the image is formed close to or on one surface, which for convenience may be called a field surface. As the image approaches a field surface, aberration coefficients tend to infinity, but it can be seen that the third degree aberrations W_{30} and W_{12} remain finite, since the infinite aberration coefficients refer to a pencil of zero aperture.

However, in certain cases, namely the transfer terms for W_{40} , W_{22} and W_{04} , and some terms in the refraction formulae for W_{22} and W_{04} , infinite values still occur, causing an indeterminacy on refraction at a surface where $t=0$ or $s=0$.

Dr. H. H. Hopkins suggested to the author that this indeterminacy might be removed by the device of computing the aberrations of a wave front at an infinite distance along the principal ray rather than at the refracting surface, and it was found that this could be carried out in the following manner: The aberrations at infinity are obtained from those at a refracting surface by putting $D = \infty$ in the transfer formulae, and they are then computed without using transfer formulae but with certain terms added to the refraction formulae; these additional terms transfer the incident wave front from infinity to the surface and transfer the refracted wave front back to infinity, and, together with those terms of the refraction formulae which become infinite, can be put into forms which are not indeterminate when $t=0$ or $s=0$. It is hoped to publish these results in detail at a later date.

(vi) *Summary of Computing Method*

To assess the state of correction of an optical system by the methods described would involve the following steps:

A paraxial ray from the axial object point is traced through the system and the data from this ray-trace are used in equations (14), (16), and (17) to determine the axial primary and secondary spherical aberration.

Several principal rays at different obliquities are traced trigonometrically or algebraically, and s - and t -fans are traced along these according to § 4(ii) to obtain the data there referred to; these data are then substituted in equations (22) and (23) to obtain the third degree aberrations, and the results of this calculation together with the s - and t -trace data are used in equations (27) to (32) to obtain the fourth degree aberrations. The final aberration coefficients are obtained as described in the respective sections.

When all the aberration coefficients have been obtained for the points where the principal rays cut the axis they may be considered in that form or they may be combined as in equation (3) of paper I, to give the aberration coefficients in polar coordinates. The latter are, perhaps more useful when the performance in all azimuths, not merely meridian and sagittal, is being considered.*

When writing down the final expression for the wave-front aberration, whether in polar or rectangular coordinates, it usually happens (if the astigmatism is reasonably low) that the extra terms in t and s which occur are negligible compared with the main terms in the aberration coefficients.†

Hence for most purposes one can write for the aberration in rectangular coordinates

$$W = N \left\{ \frac{1}{2} \left(\frac{1}{t} - \frac{1}{R} \right) y^2 + \frac{1}{2} \left(\frac{1}{s} - \frac{1}{R} \right) z^2 + C_{30} y^3 + C_{12} y z^2 + C_{40} y^4 + C_{22} y^2 z^2 + C_{04} z^4 + O(\rho^5) \right\}, \quad (33)$$

* It seems to be impossible to obtain direct refraction formulae for the polar aberration coefficients, as may be seen, for example, by attempting to combine (13) and (28) of paper I according to (3) of paper I to obtain a refraction formula for B_{33} ; it will be found that the angle of incidence cannot be eliminated from the left-hand side of the equation.

† In the example given in § 5 below the neglect of these terms caused an error of less than 0.1 wavelength at any point in the pupil for the pencil at 15° field angle; the error was less than 0.02 wavelength for the pencil at 12.5° field angle. The astigmatic differences in these two cases were about 0.7 inch and 0.2 inch respectively.

and in polar coordinates

$$\begin{aligned}
 W = N \left\{ \frac{1}{2} \left(\frac{1}{t} - \frac{1}{s} \right) \rho^2 \cos^2 \phi + \frac{1}{2} \left(\frac{1}{s} - \frac{1}{R} \right) \rho^2 + (C_{30} - C_{12}) \rho^3 \cos^3 \phi \right. \\
 + C_{12} \rho^3 \cos \phi + (C_{40} - C_{22} + C_{04}) \rho^4 \cos^4 \phi \\
 \left. + (C_{22} - 2C_{04}) \rho^4 \cos^2 \phi + C_{04} \rho^4 + O(\rho^5) \right\}, \quad \dots (34)
 \end{aligned}$$

where R is the radius of the reference sphere, i.e. the distance along the principal ray from the wave front to the plane of focus considered. The values of the C 's may refer either to the point where the principal ray leaves the last surface or to the point where it cuts the axis, and t and s should be chosen accordingly.

Four significant figures are sufficient in the aberration calculations, and in practice it often happens that only three or two figures of a number of terms are significant. In particular the contributions of the transfer terms to the fourth degree coefficients are often very small unless there is much uncorrected third degree aberration in a long air space.

The above procedure gives all the aberrations which affect the imagery of off-axis object points except distortion, which is obviously obtained directly from the principal ray-trace, and field curvature. Field curvature might be defined as the part of the wave-front aberration varying as $\rho^2 f(\sigma^2)$ (where $f(\sigma^2)$ is an even function of σ , the field angle) which is independent of astigmatism; this would define a generalized Petzval surface which would be the image surface in the absence of astigmatism. Although this definition is valid for primary aberrations, it is questionable whether it has a meaning when fourth and higher powers of σ are considered, since it is not obvious that an expansion for astigmatism involving σ^4 etc. would contain terms independent of the angles of incidence of the principal ray. However, the shapes of the s and t focal surfaces are more important for assessing the quality of the image, and these are known from the calculations. The choice of the radius of the reference sphere then enables one to consider the imagery on any focal surface, plane or otherwise.

§ 5. CLOSENESS OF APPROXIMATION OF THE FORMULAE

The formulae give the terms up to and including those of the fourth degree in the aperture in the expansion of the wave-front aberration as a power series, and it is desirable to determine the errors which occur through neglecting the remaining terms. The approach chosen was to investigate numerically systems of the types for which such a computing scheme is likely to be useful—systems of moderate aperture with fairly large angles of incidence of the principal ray. Several meridian and skew rays of the pencils surrounding the principal rays were traced through such a system and the resulting transverse aberrations were plotted and integrated graphically to determine the *true* wave-front aberrations; these were compared with the power series results. Some of the results of this comparison and a brief description of the method will be given here.

The work involved much skew ray tracing, and after several trials it was decided that the most convenient skew tracing method to use would be that described by Smith (1923); this method requires no trigonometrical tables, has no special cases where significant figures are lost, and can be checked completely at every stage.

To obtain the numerical integration formulae, let the equation of a wave front be equation (2) of paper I, and let (x, ρ, ϕ) be cylindrical polar coordinates of the point where a ray meets the wave front, so that $y = \rho \cos \phi$, $z = \rho \sin \phi$. Let (R, Y, Z) be the rectangular coordinates of the intersection of the ray with an image plane whose equation is $x = R$. Then by geometry

$$Y = \rho \cos \phi - \left\{ \frac{\partial x}{\partial \rho} \cos \phi - \frac{1}{\rho} \frac{\partial x}{\partial \phi} \sin \phi \right\} (R - x) + O(\rho^4),$$

$$Z = \rho \sin \phi - \left\{ \frac{\partial x}{\partial \rho} \sin \phi + \frac{1}{\rho} \frac{\partial x}{\partial \phi} \cos \phi \right\} (R - x) + O(\rho^4),$$

so that $Y \cos \phi + Z \sin \phi = \rho - (R - x) \frac{\partial x}{\partial \rho} + O(\rho^4)$.

On integrating this with respect to ρ , substituting the value of x in polar coordinates, and re-grouping the terms, one obtains

$$\begin{aligned} -\frac{1}{R} \int_0^e \{Y \cos \phi + Z \sin \phi\} d\rho &+ \frac{1}{8R} \left\{ \left(\frac{1}{t} - \frac{1}{s} \right) \cos^2 \phi + \frac{1}{s} \right\}^2 \rho^4 - \frac{1}{8R^3} \rho^4 \\ &= \frac{1}{2} \left(\frac{1}{t} - \frac{1}{s} \right) \rho^2 \cos^2 \phi + \frac{1}{2} \left(\frac{1}{s} - \frac{1}{R} \right) \rho^2 + (C_{30} - C_{12}) \rho^3 \cos^3 \phi + C_{12} \rho^3 \cos \phi \\ &\quad + \left\{ C_{40} - C_{22} + C_{04} + \frac{1}{8} \left(\frac{1}{t^3} - \frac{2}{t^2 s} + \frac{1}{s^3} \right) \right\} \rho^4 \cos^4 \phi \\ &\quad + \left\{ C_{22} - 2C_{04} + \frac{1}{4s} \left(\frac{1}{t^2} - \frac{1}{s^2} \right) \right\} \rho^4 \cos^2 \phi + \left\{ C_{04} + \frac{1}{8} \left(\frac{1}{s^3} - \frac{1}{R^3} \right) \right\} \rho^4 + O(\rho^5). \end{aligned}$$

.....(35)

This equation is arranged so that the right-hand side gives the wave-front aberration terms up to those in ρ^4 , plus the remainder term, while the left-hand side gives the *true* wave-front aberration of all orders; the first term on the left-hand side is the integral of the transverse aberration referred to above. In practice several terms in equation (35) are usually negligible, as explained at the end of §4, and it then takes the simpler form

$$\begin{aligned} -\frac{1}{R} \int_0^e (Y \cos \phi + Z \sin \phi) d\rho &= \frac{1}{2} \left(\frac{1}{t} - \frac{1}{s} \right) \rho^2 \cos^2 \phi + \frac{1}{2} \left(\frac{1}{s} - \frac{1}{R} \right) \rho^2 \\ &\quad + (C_{30} - C_{12}) \rho^3 \cos^3 \phi + C_{12} \rho^3 \cos \phi + (C_{40} - C_{22} + C_{04}) \rho^4 \cos^4 \phi \\ &\quad + (C_{22} - 2C_{04}) \rho^4 \cos^2 \phi + C_{04} \rho^4 + O(\rho^5). \end{aligned}$$

.....(36)

To use this equation three or four rays are traced at each value of ϕ to be considered, Y and Z are determined for each ray, and the integral on the left-hand side is evaluated graphically and tabulated as a function of ρ . This is compared with the right-hand side, also considered as a function of ρ , and the difference gives the remainder terms of order ρ^5 and higher which should occur in the power series for the wave-front aberration. In the examples to be given, the axial secondary spherical aberration was added to the right-hand side, to improve the fit.

The results to be given here were from calculations on a design of a 20-inch F/6.3 triplet objective of which the specification was kindly made available by Messrs Kodak Ltd. The objective was designed to work up to 10° semi-field angle, but the calculations were made for angles up to 15° , since the aberrations were still of a suitable magnitude for this purpose.

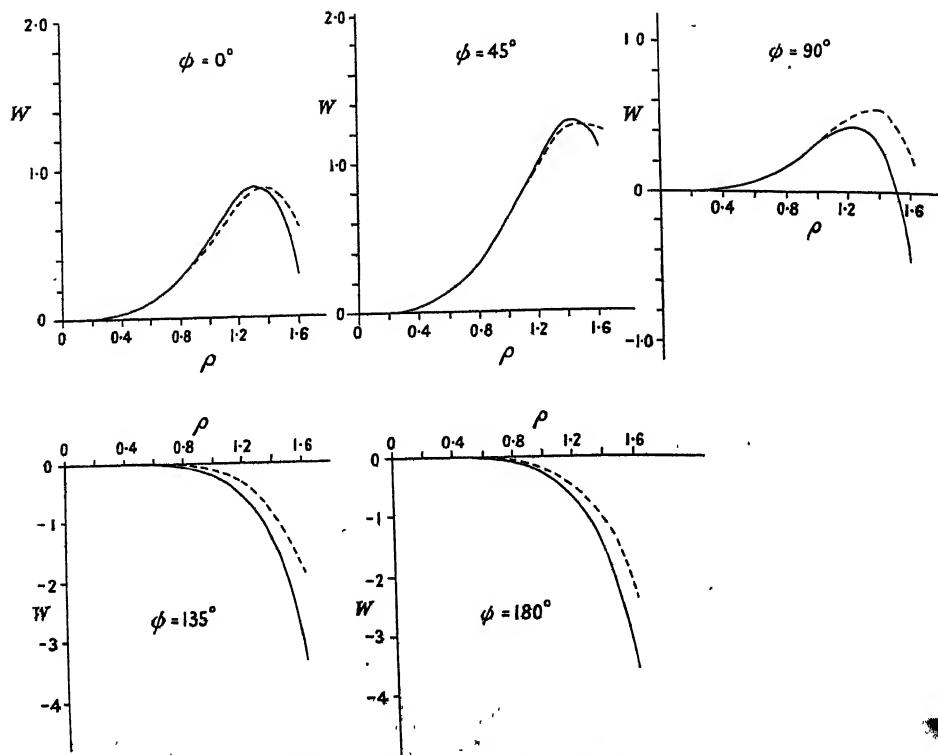


Figure 1. Wave-front aberration at $U_{pr} = -70^\circ$.
Aberration W in microns; semi-aperture ρ in inches.
Full lines: from ray tracing. Dotted lines: from aberration coefficients.

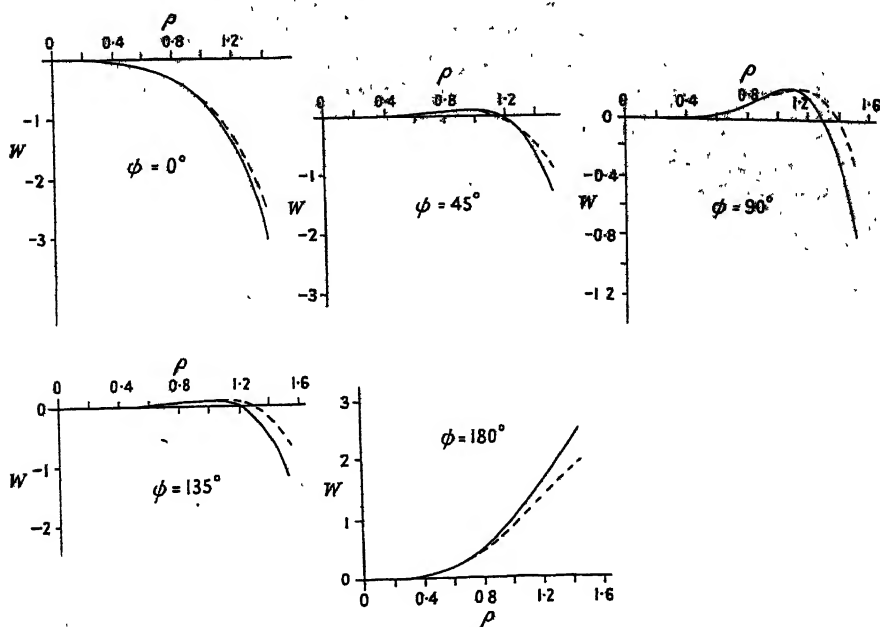


Figure 2. Wave-front aberration at $U_{pr} = -10^\circ$.
Aberration W in microns; semi-aperture ρ in inches.
Full lines: from ray tracing. Dotted lines: from aberration coefficients.

The aberrations are given in microns of optical path and the aberration coefficients are in microns per inch³ and microns per inch⁴ for the third and fourth degree coefficients respectively. The coefficients for rectangular coordinates, referred to the exit pupil, are tabulated against U_{pr}' , the convergence angle of the emerging principal ray, and these can be directly substituted in equations (33)

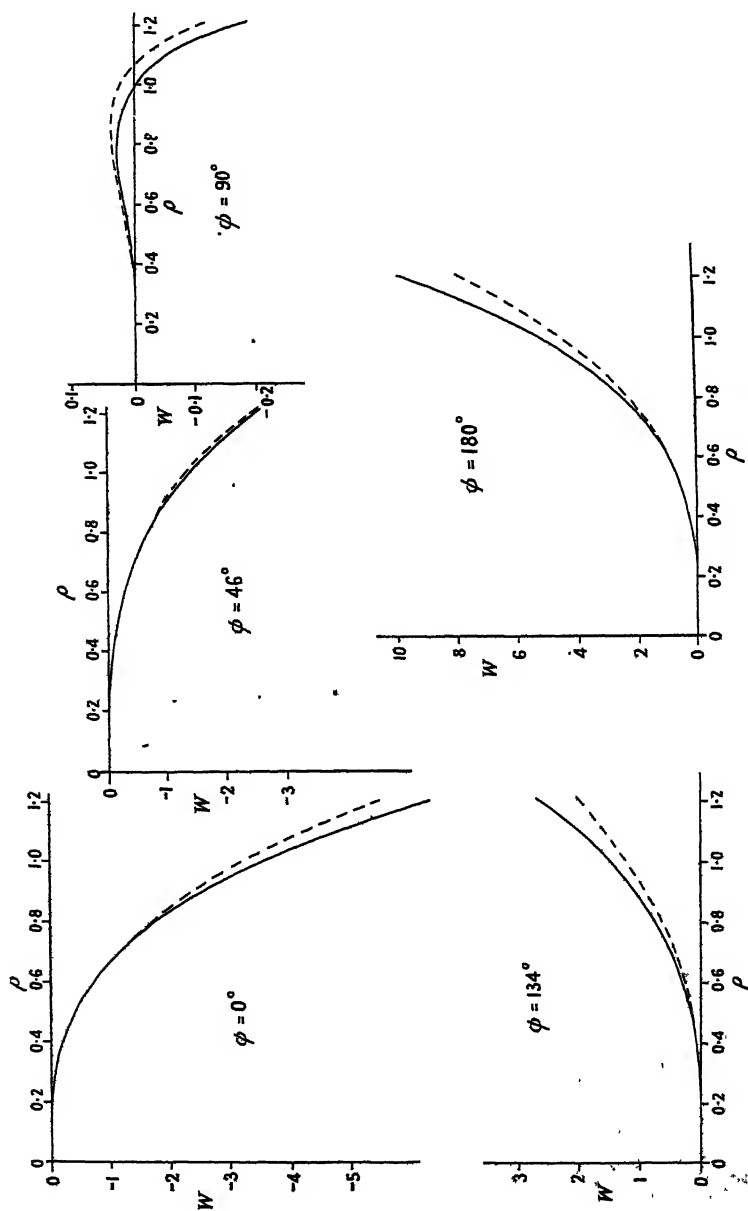


Figure 3. Wave-front aberration at $U_{pr} = -12\frac{1}{2}^\circ$. Aberration W in microns, semi-aperture ρ in inches. Full lines: from ray tracing. Dotted lines: from aberration coefficients.

and (34); the results of substituting in equation (34) are given below the table. In the graphs W , the aberration as found from the coefficients (broken lines) and by ray-tracing (full lines) is plotted against ρ , the semi-aperture, for various azimuths at each value of U_{pr} (Figures 1 to 4), and for an axial pencil (Figure 8). The

reference sphere in each case is chosen to give no second degree aberration (error of focus) in the azimuth concerned. The s - and t -fields and the polar aberration coefficients are plotted against U_{pr} in Figures 5, 6 and 7.

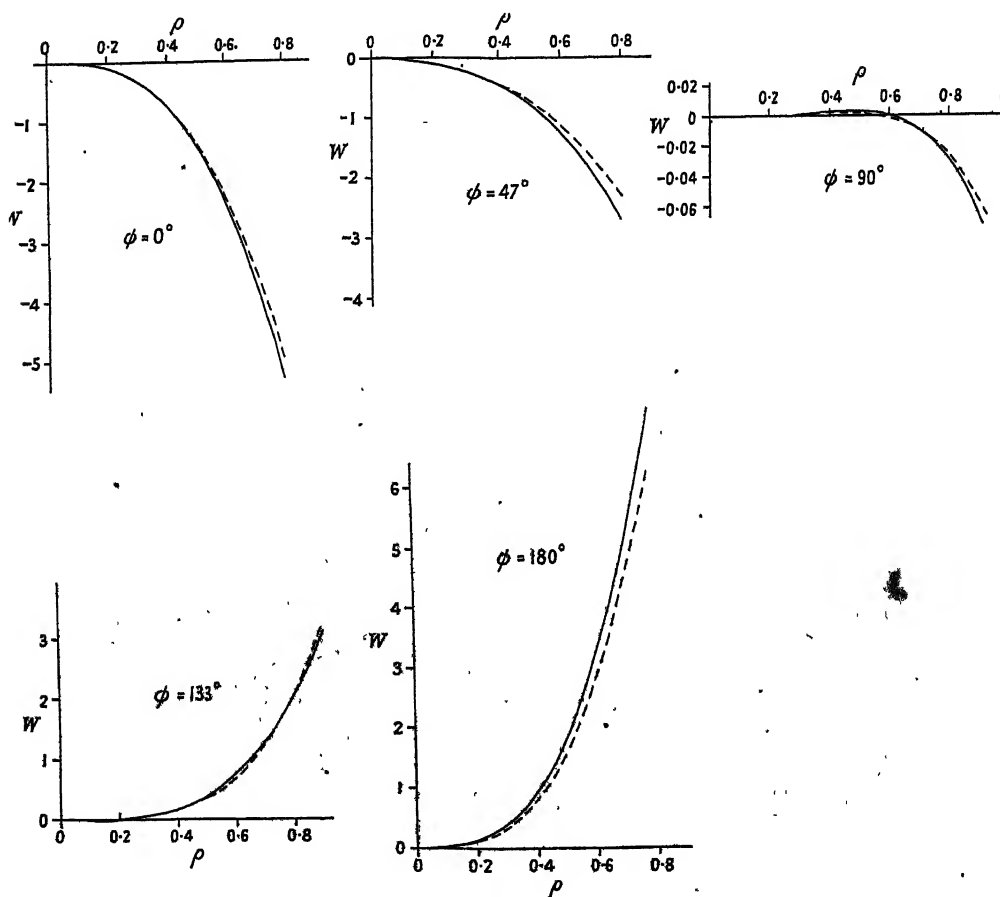


Figure 4. Wave-front aberration at $U_{pr} = -15^\circ$.
Aberration W in microns, semi-aperture ρ in inches.
Full lines: from ray tracing. Dotted lines: from aberration coefficients.

Aberration Coefficients of a Triplet Objective

U_{pr} (degrees)	U_{pr}'	C_{12} (microns/inch ³)	C_{30}	C_{40}	C_{22} (microns/inch ⁴)	C_{04}
0	0	0	0	0.649	1.298	0.649
-7	-7.0237	0.713	0.350	0.325	0.856	0.497
-10	-10.0406	0.658	-0.778	0.272	0.425	0.345
-12.5	-12.5655	0.056	3.85	0.858	-0.150	0.201
-15	-15.1126	-1.700	-11.95	2.77	3.33	0.028

Axial secondary spherical aberration: $C_s = -0.177$ microns/inch³.

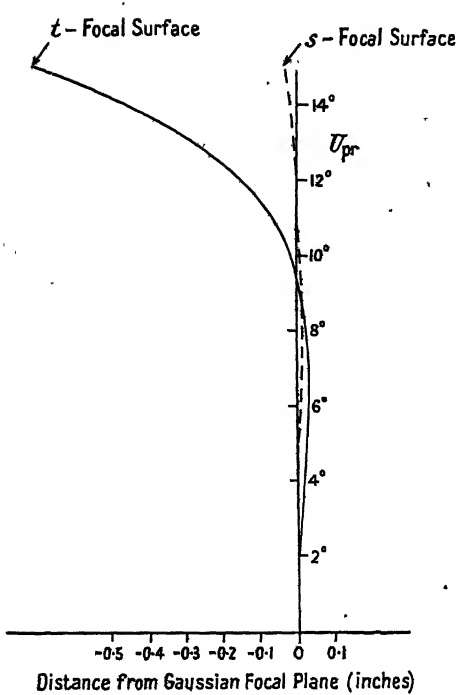
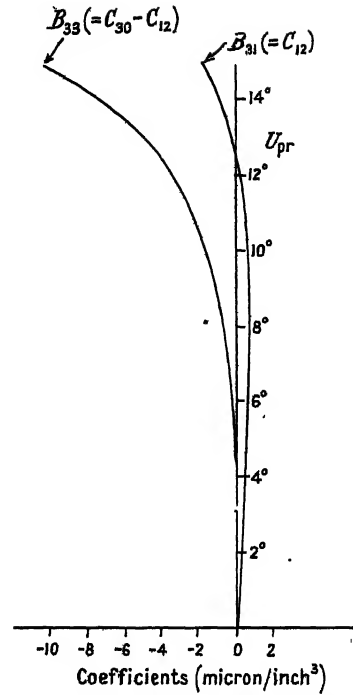

 Figure 5. *s*- and *t*-focal surfaces.


Figure 6. Third degree polar aberration coefficients.

Wave-front aberration in polar co-ordinates (from equation (34)), with reference sphere centred on the sagittal focus:

U_{pr} (degrees)	W (microns)
-7	$\begin{cases} -0.808\rho^2\cos^2\phi - 0.363\rho^3\cos^3\phi + 0.713\rho^3\cos\phi - 0.034\rho^4\cos^4\phi \\ -0.138\rho^4\cos^2\phi + 0.497\rho^4 - 0.177\rho^6 \end{cases}$
-10	$\begin{cases} 0.615\rho^2\cos^2\phi - 1.436\rho^3\cos^3\phi + 0.658\rho^3\cos\phi + 0.196\rho^4\cos^4\phi \\ -0.273\rho^4\cos^2\phi + 0.349\rho^4 - 0.177\rho^6 \end{cases}$
-12.5	$\begin{cases} 5.95\rho^2\cos^2\phi - 3.91\rho^3\cos^3\phi + 0.056\rho^3\cos\phi + 1.209\rho^4\cos^4\phi \\ -0.552\rho^4\cos^2\phi + 0.201\rho^4 - 0.177\rho^6 \end{cases}$
-15	$\begin{cases} 20.3\rho^2\cos^2\phi - 10.25\rho^3\cos^3\phi - 1.70\rho^3\cos\phi - 0.53\rho^4\cos^4\phi \\ + 3.39\rho^4\cos^2\phi + 0.03\rho^4 - 0.177\rho^6 \end{cases}$

§ 6. DISCUSSION OF THE RESULTS

The purpose of analytical aberration calculations is to obtain a useful approximation to the aberrations, with an indication of the relative contributions of the various components of the system. It is immaterial in most cases if the *final* result is in error by say, 10%, since tolerances cannot be assigned with such precision. An exact knowledge of the aberrations, such as might be obtained by extensive ray tracing, would add nothing of further value for assessing the general

performance and would give no indication as to the contributions of individual surfaces. From this point of view it can be seen from the graphs that a useful approximation to the wave-front aberration is obtained for the aperture used. The angles of incidence of the principal rays ranged up to 26° on the air side of a glass-air surface.

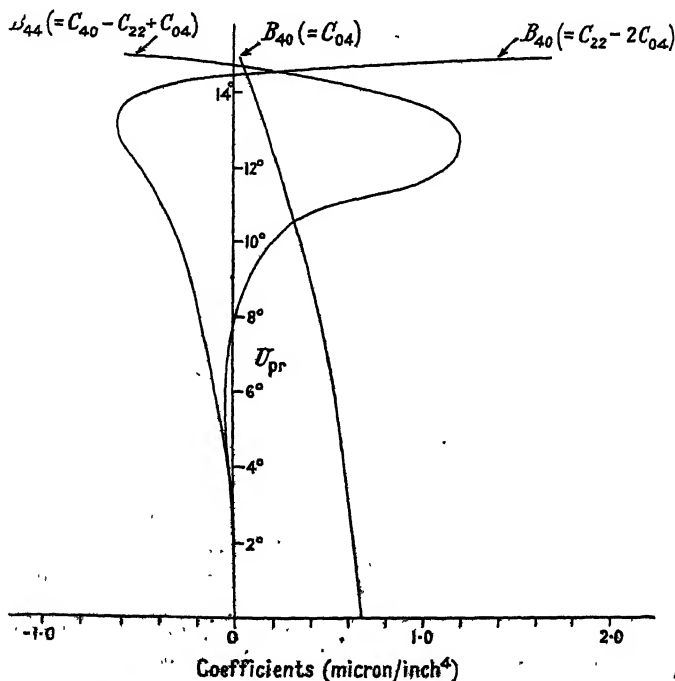


Figure 7. Fourth degree polar aberration coefficients.

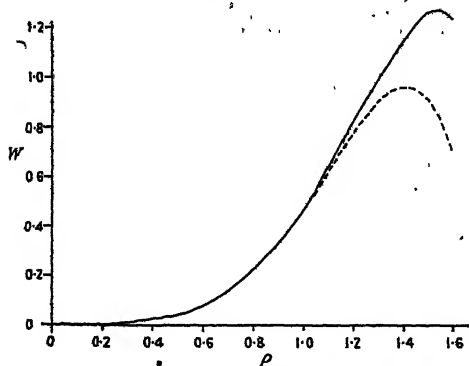


Figure 8. Axial aberration.

Aberration W in microns; semi-aperture ρ in inches.
Dotted line: from ray tracing. Full line: from aberration coefficients.

Some interesting points emerge from a study of the graphs showing the variations of the coefficients with field angle. It can be seen that C_{30} is zero when U_{pr} is $-8\frac{1}{2}^\circ$, so that from meridian ray-traces one would draw the conclusion that there was no coma at this field angle; however, the value of C_{13} shows that at 45° azimuth in the pupil there is about 0.5μ of coma. Again, at $U_{pr} = -9^\circ$

the value of C_{40} shows about 0.2μ of spherical aberration in the meridian plane, but there is 0.4μ in the sagittal plane and the run of the coefficients shows that it is almost 0.4μ at nearly all azimuths out of the meridian plane. On the other hand, at -15° there is about 3μ of spherical aberration in the meridian plane, but this rapidly decreases to nearly zero in the sagittal plane.

In Figure 6 the shape of the B_{31} graph (coma varying in the aperture as $\rho^3 \cos \phi$) shows that in addition to the Seidel component, which varies linearly with the field angle U_{pr} there is a component of opposite sign varying as U_{pr}^3 (an aberration of type $\rho^3 \sigma^3 \cos \phi$ in the usual aberration classification mentioned in I).

Similarly the B_{33} graph (coma varying as $\rho^3 \cos^3 \phi$) contains components varying as U_{pr}^3 , the lowest power which could occur, and as U_{pr}^5 .

In Figure 7 the ordinary spherical aberration coefficient (B_{40} , for aberration varying as ρ^4) shows, in addition to the constant axial component, a part varying as U_{pr}^2 ; B_{42} , the coefficient of the aberration varying as $\rho^4 \cos^2 \phi$, shows components varying as U_{pr}^2 and U_{pr}^4 , while B_{44} has components varying as U_{pr}^4 and U_{pr}^6 .

The general conclusion may be drawn that for systems of moderate aperture and extended field the use of this method of assessment will show up properties which could not otherwise have been discovered without a large amount of ray-tracing. An estimate based on the time taken for the above computations suggests that to determine the final wave-front aberrations by ray-tracing would take about four times as long as the calculation of the aberration coefficients. An additional advantage is that the aberration calculations give immediately the contributions of each part of the system to the final aberrations.

ACKNOWLEDGMENTS

This work is published by permission of the Chief Scientist, the Ministry of Supply. Thanks are due to Professor L. C. Martin and Dr. H. H. Hopkins for advice and criticism, and to Miss J. M. Drewitt, who did the computations.

REFERENCES

- CONRADY, A. E., 1929, *Applied Optics and Optical Design*, Part I (Oxford : University Press).
 HOPKINS, H. H., 1946, *Proc. Phys. Soc.*, **58**, 663.
 VON ROHR, M., 1920, *The Formation of Images in Optical Instruments*, translated by R. Kanthack (London : His Majesty's Stationery Office).
 SMITH, T., 1923, *Dictionary of Applied Physics*, Vol. IV (London : Macmillan and Co. Ltd.), article on "Optical Calculations".
 WEINSTEIN, W., 1949, *Proc. Phys. Soc. B*, **62**, 726.

LETTERS TO THE EDITOR

An Optical Method for Studying the Deformation of Mild Steel

The purpose of this note is to indicate some of the work which has continued over the past few years at the Cavendish Laboratory on the Lüders' deformation of mild steel. In particular, a new 'optical strain gauge' method will be described, which has been used in studying the properties of strain-aged materials.

When a normal tensile test-piece of mild steel is strained, the stress-strain curve which results is very jagged, because of the generation of a complex system of Lüders' bands. Some years back, however, Dr. E. Orowan and Mr. W. Chitruk found that this difficulty could be avoided by the use of thin wire specimens. The Lüders' band is nucleated at one grip, after a marked upper yield point, and spreads over the length of the specimen under virtually constant stress. The investigation was carried a stage further here by Mr. Sylwestrowicz in work for the British Iron and Steel Research Association. He made a preliminary study of the yield point and strain-ageing properties of the material, and found that, if a Lüders' band is propagated along a specimen, the load released, and the specimen aged for short times in the temperature interval 50°C. – 200°C. , a new upper yield point appears before the band spreads further. After longer ageing times, however, the front of the band is so securely locked that a new band appears from the grip, propagating in the *old* one. This new band will be called the 'secondary' Lüders' band, and the original the 'primary' band. The propagation stress for the secondary band may be anything up to 40% more than the primary one.

In the course of a systematic study by the writer of the strain-ageing properties of these wire specimens, it was necessary to devise some method which would allow the development of the secondary band to be followed. A technique has been adopted which was originally used in a study of the jerky extension of zinc crystals by Orowan (1949). Fine filings of tin are dropped down a tube of molten stearic acid. The heating arrangements are such that the temperature at the top of the tube is above the melting point of tin (232°C.) and below it at the bottom. Thus, the tin filings melt, and solidify as balls at the bottom. The molten stearic acid is then poured off and, when cool, the balls are extracted and cleaned with ether.

The balls are mounted half-way down the wire tensile specimen on a thin thread of rubber solution, and are illuminated by a Western 'concentrated arc' lamp. The images of the source in the balls are focused on to a strip of photographic paper in a drum camera. A good lens of short focal length must be used, giving an optical magnification of $\times 15$ to $\times 50$.

Typical records are shown diagrammatically in Figure 1. In the extensometer, the bottom end of the specimen is pulled down, and the consequent traces on the record are the inverted pictures of the motion of each ball. The start of each trace is under zero stress. As the load comes on, the balls move downwards with a velocity of $V/2$. When a band is nucleated at the bottom grip, as shown in Figure 1(a), no deformation occurs outside the band, the balls remaining fixed in space, and the trace is horizontal. As the edge of the band moves past the balls, all the deformation occurs above them, and they move downward with the full velocity of the straining head. Thus, the gradient of the trace is twice that in the elastic region. The line of transition is staggered across the record, as shown by the hypotenuse of the dotted triangle in the diagram. When the band has spread to the top grip, deformation is from then on homogeneous, and the gradient of the trace is the same as in the elastic region.

Figure 1(b) shows the traces in the similar case when the band is nucleated in the top grip, while Figure 1(c) shows the more rare example of two bands generated simultaneously.

Figure 2(a) shows the photograph of a typical record, obtained from undeformed material, and similar in form to Figure 1(a) above. From the gradients of the traces, and knowing the velocity V of the straining head and the rate of rotation of the camera drum, we may determine the optical magnification, the rate of advance of the band edge, and the Lüders' strain. For this particular record, the values obtained are $\times 18$, 0.76 cm/min. , and 4.3% respectively.

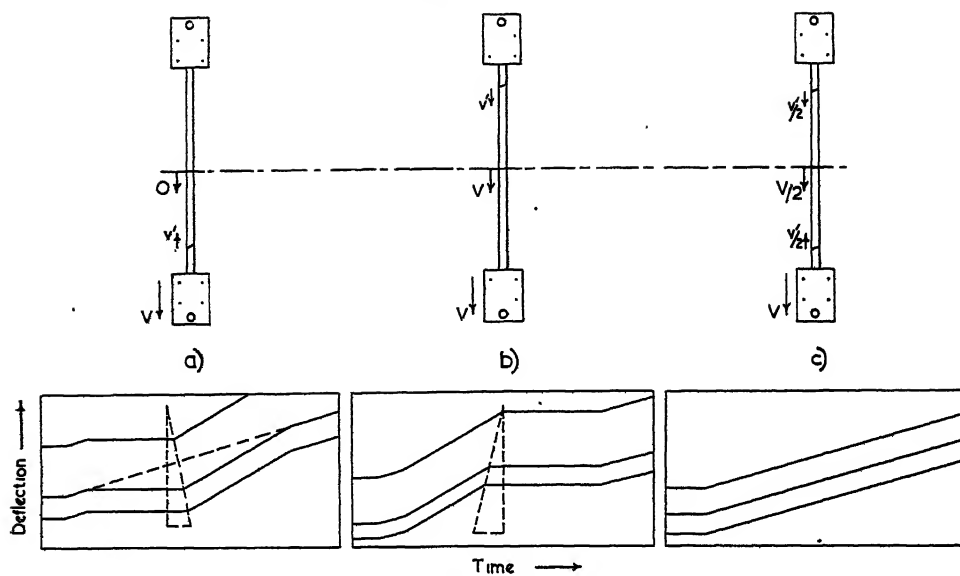


Figure 1. Diagram of typical records.

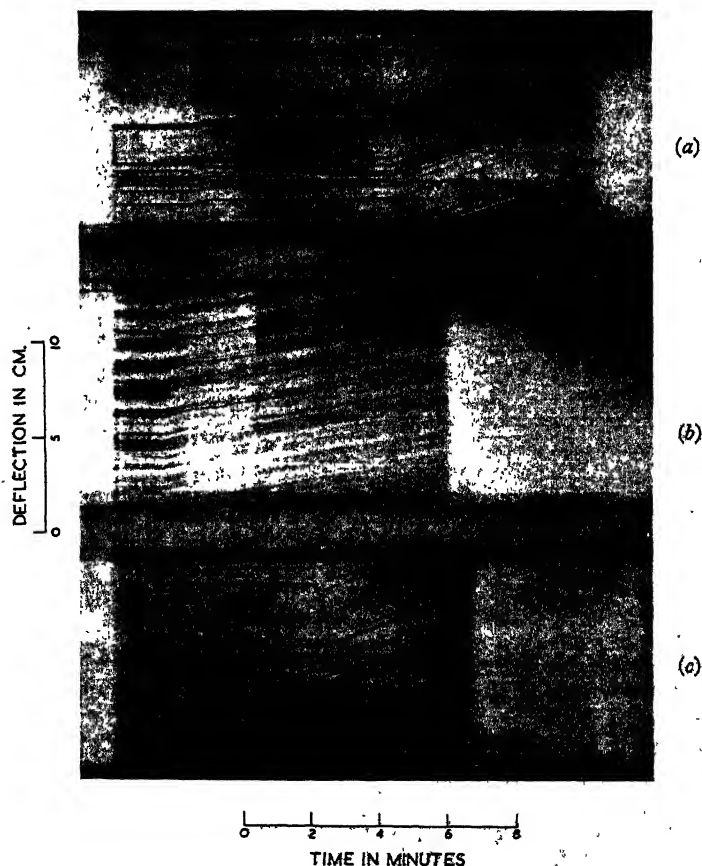


Figure 2. Photograph of records.

If the specimen is strained to the end of the Lüders' extension, and then aged, the secondary band slowly appears. Figure 2 (b) shows the deformation after $4\frac{1}{2}\%$ plastic extension and ageing for 30 minutes at 60°C . Here, a band has started from the lower grip, but material outside the band is deforming simultaneously, perhaps as localized bands. Soon the band edge meets material deformed to the same order of plastic strain as its own Lüders' strain, and so the band spreads no further. Deformation then spreads from another point. All this occurs under constant stress.

Figure 2 (c) shows a later stage of the secondary band development, after $4\frac{1}{2}\%$ strain and one hour at 60°C . The record is now similar to (a), but even here a little deformation occurs outside the band, and the traces are never quite horizontal.

The mechanics of this transition type of deformation are still under investigation.

One other important result has arisen from this work. It was found that if the grain size of the specimen is larger than a certain critical value, Lüders' deformation is no longer possible. For a basic Bessemer steel, containing 0.045% C, this value is 60 ± 3 grains/mm. For coarser grained specimens, deformation is still under constant stress, but the traces obtained show that the actual mechanism involved is similar to the transition type obtained above. No amount of straining or ageing will ever induce a Lüders' band to appear if the grain size is above this critical limit.

Cavendish Laboratory,
Cambridge.
5th June, 1950.

E. O. HALL.

OROWAN, E., 1949, *J. Metals (Metals Trans.)*, **185**, 876

The Scattering of 3-cm. Radiation by Ionized Gases

Work has been carried out in this laboratory on ultra-high-frequency techniques for studying electron concentrations in discharges, particularly on the scattering of radiation from low pressure discharges (cf. the meteor work, especially the theoretical work of Herlofson (1948 a, b)) for wavelengths of 3 cm., but also on the absorption in discharges of such radiations (see Andrew *et al.* 1948, Adler 1949 for similar recent work on absorption). Since, so far as we know, no other published account of scattering from laboratory discharges has been given, although experiments on the aurora have been reported (Lovell *et al.* 1947, Forsyth *et al.* 1950), it was thought worth while to report the following preliminary experiments. The experiments on absorption will be reported elsewhere.

Experiments have been made with two sealed-off commercial mercury lamps of diameters approximately 3.1 cm. (17 cm. long) and 6.4 cm. (40 cm. long), and the experimental arrangement comprises transmitting and receiving horns arranged prependicular to each other so that scattered radiation from the discharge is measured, using a crystal matched to the receiving waveguide (VSWR 1.08). Figures 1 and 2 show the scattered power in arbitrary units as a function of the tube currents (note that the Hg vapour pressure varies, because of temperature variations, with the current) for two directions of the electric vector (perpendicular and parallel to the discharge axis).

The scattered radiation from the glass tube, at zero current, is particularly important in the case of Figure 1 and the possibility of interference between the radiation scattered from the plasma and from the front and rear walls of the tube for all cases is an obvious factor.

If the effects of magnetic fields, both strays and the self-field of the discharge, are neglected and if collision damping is also negligible, the electron concentration n may be found, if the scattering shows a sharply defined onset and is equivalent to the ionospheric refraction, by the simple critical frequency formula (dielectric constant $\epsilon=0$)

$$n_{\text{crit}} = 1.24 \times 10^{-8} f^2;$$

f is the radiation frequency in cycles per second. For $\lambda=3$ cm., $n_{\text{crit}}=10^{12}/\text{cm}^3$, and this is not unreasonable as a maximum axial electron concentration for the discharges considered (approximately 5% ionization) at the currents shown in Figures 1 and 2. This treatment is, no doubt, an over-simplification and the following comments are pertinent.

1. The three cases of radiation scattering (wavelength λ), for discharges of diameter d are: (a) extended scatterer of dimensions very much greater than d (ionosphere, see for example, Mitra 1949); (b) cylindrical discharge, $d \ll \lambda$ (meteor trails), where the scattering may be calculated as shown by Blackett and Lovell (1941), Lovell and Clegg (1948) and Lovell (1950) if ϵ is of the order unity; (c) cylindrical discharge, d , of the order of λ (present discharges) where the radiation, scattered across a diameter, shows phase differences (Herlofson 1948) if ϵ is of the order unity.

2. Herlofson states that when ϵ is approximately zero (present discharges) at the centre of the scattering column, the scattering with the electric vector parallel to the discharge column can be calculated as in cases (1 b) and (1 c) up to 10^{18} electrons per centimetre

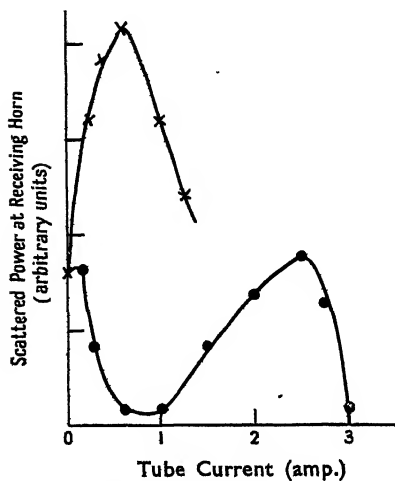


Figure 1. Small tube.

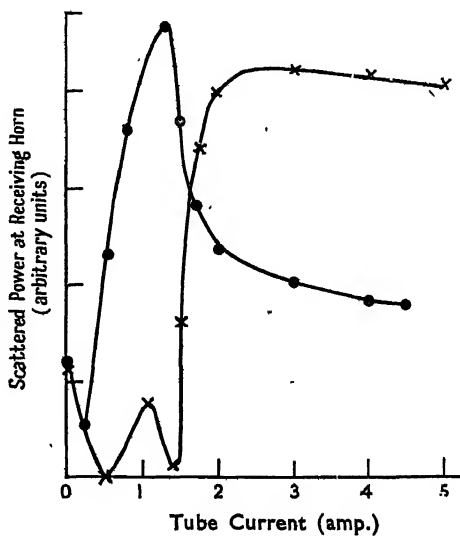


Figure 2. Large tube.

● E vector parallel to tube axis.
 × E vector perpendicular to tube axis.

Units on the ordinate axis for the two curves are not comparable since the tube dimensions are different, but the curves are normalized at zero current.

length of scatterer and may be much less than when the electric vector is perpendicular to the discharge column (perhaps by several powers of 10), since in the latter case plasma oscillations may be excited. This type of effect seems to be apparent with the small tube (Figure 1) but not with the large tube (Figure 2).

The experiments are being extended in various directions and will be reported in more detail, with the absorption work, elsewhere. It is intended to compare the electron concentration measurements with those obtained with probe and spectroscopic techniques.

Department of Electrical Engineering,
 The University of Liverpool.
 16th April 1950, in final form 3rd July 1950.

S. N. DENNO.
 H. A. PRIME.
 J. D. CRAGGS.

- ADLER, F. P., 1949, *J. Appl. Phys.*, **20**, 1125.
 ANDREW, E. R., AKFORD, D. W. E., and SUGDEN, T. M., 1948, *Trans. Faraday Soc.*, **44**, 427.
 BLACKETT, P. M. S., and LOVELL, A. C. B., 1941, *Proc. Roy. Soc. A*, **177**, 183.
 FORSYTH, P. A., PETRIE, W., VAWTER, F., and CURRIE, B. W., 1950, *Nature, Lond.*, **165**, 361.
 HERLOFSON, N., 1948 a, *Observatory*, **68**, 230; 1948 b, *Rep. Prog. Phys.*, **11**, 444 (London: Physical Society).
 LOVELL, A. C. B., 1950, *Sci. Prog.*, **38**, 22, and references there cited.
 LOVELL, A. C. B., and CLEGG, J. A., 1948, *Proc. Phys. Soc.*, **60**, 491.
 LOVELL, A. C. B., CLEGG, J. A., and ELLYETT, C. D., 1947, *Nature, Lond.*, **160**, 372.
 MITRA, S. K., 1949, *The Upper Atmosphere* (Calcutta).

Temperature Dependence of Counter Characteristics in Self-Quenching Geiger-Müller Counters.

Parkash and Kapur in a recent paper (1950) have reported on the variation of characteristics with temperature of the usual type of argon-alcohol counter. They used counters of two types, viz. those in internal copper cathodes and those with external graphite cathodes. They noted that in counters with internal cathodes, the plateau decreased in length and disappeared at about 9° C., but that no such effect was obtained with the second type of counter. At higher temperatures (up to 62° C.) they obtained an increase of plateau slope and a decrease of plateau length in both cases.

Our experience at Harwell concerning the effect of increase of temperature on plateau slope and plateau length confirms the results of Parkash and Kapur, but we do not agree

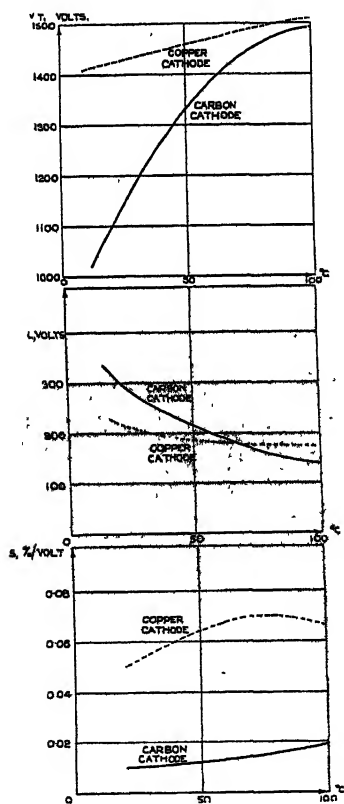


Figure 1. Variation of threshold voltage (V_T), plateau length (L) and plateau slope (S) with change of temperature for a counter with argon-alcohol filling.
 — internal metal cathode counter.
 --- internal graphite cathode counter.

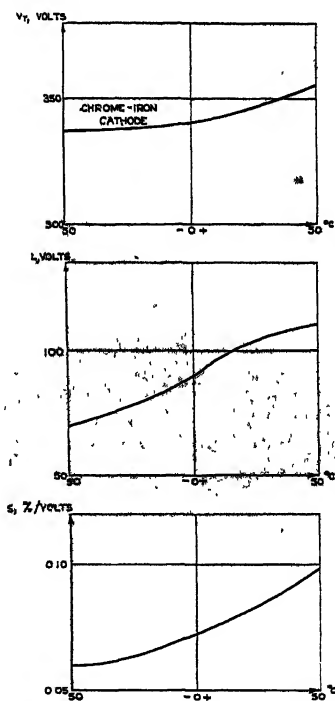


Figure 2. Variation of threshold voltage (V_T), plateau length (L) and plateau slope (S) with change of temperature for an internal metal cathode counter with an argon-neon-bromine filling gas.

with the deductions they make about the advantage of the external cathode counter ('Maze' tube) at low temperatures.

Typical results, obtained at Harwell, showing the variation of threshold voltage V_T , plateau length L and plateau slope S with change of temperature are shown in Figure 1. Two types of counters were used for these tests, viz. (i) a counter with an internal metal cathode and (ii) a counter with an internal graphite cathode. The filling gas was the argon-alcohol mixture as used by Parkash and Kapur and the change of characteristics with

increase of temperature is as reported by these authors. It will be noted that the temperature coefficient, defined as the average change of threshold voltage per degree Centigrade, is higher for the graphite cathode counter than for the metal cathode counter. This is to be expected, since the adsorption of alcohol vapour by a graphite surface is very temperature dependent.

Tests have also been made at Harwell using the Maze (1946) type of counter with an external cathode on the outside of a glass envelope. The temperature coefficient obtained was very low as compared with that obtained with the internal cathode type of counter, as also reported by Parkash and Kapur, but in our tests it was noted that the performance was markedly affected by the electrical resistance of the glass. In practice, the electrical resistance of the glass becomes too high at low temperatures to allow this type of counter to operate. For this reason we do not agree with the conclusion of Parkash and Kapur that the Maze type of counter with the usual argon-alcohol filling is satisfactory for low temperature operation. The Maze type of counter has several advantages, e.g. simple construction, great robustness (excess voltages may be used without destroying the characteristics), etc., but like all counters using ethyl alcohol as the quenching agent it suffers from the disadvantage that the alcohol condenses at low temperatures. In our view, if low temperature operation is required, it is better to use a counter with an internal cathode, but with a quenching agent other than ethyl alcohol. When operation at temperatures of the order of 5° C. or below has been a requirement, we have used ethyl bromide as the quenching agent, but more recently we have used bromine, and now all our Geiger counter instruments for geological prospecting (for which operation over a wide temperature range is required) are using counters with this quenching agent. Some typical results for a counter of this type are shown in Figure 2. Counters with bromine as the quenching agent have other advantages, as has already been reported (Franklin and Loosemore 1950, Taylor 1950). Of these other advantages, probably their long life ($\sim 10^{10}$ counts) is the most important.

Atomic Energy Research Establishment,
Harwell, Berkshire.
3rd July 1950.

W. R. LOOSEMORE.
DENIS TAYLOR.

FRANKLIN, E., and LOOSEMORE, W. R., 1950, *J. Instn. Elect. Engrs.*, in press.
MAZE, R., 1946, *J. Phys. Radium*, 7, 164.
PARKASH, O., and KAPUR, P. L., 1950, *Proc. Phys. Soc. B*, 63, 461.
TAYLOR, D., 1950, *J. Sci. Instrum.*, 27, 81.

REVIEWS OF BOOKS

Some Recent Researches in Solar Physics, by F. HOYLE. Pp. xii+134.
Cambridge Monographs on Physics. 1st Edition. (Cambridge:
University Press, 1949). 12s. 6d.

Mr. Hoyle has given us a revolutionary little book which will have a profoundly stimulating effect upon the development of solar physics.

Without prejudice to the splendid contributions of the pioneer workers of fifty years ago, such as Lockyer, Evershed and Hale, we may say that the present situation in solar physics is the outcome of the development of new experimental techniques developed within the last twenty years. The author has been quick to accept the challenge presented by this accumulation of new knowledge, and he has worked out in detail an entirely novel approach to the main problem. This is the *accretion hypothesis* which he has developed jointly with Lyttleton and Bondi.

The central problem may be briefly stated as follows. All our preconceived ideas would lead us to suppose that the sun's outer atmospheric layers are cooler than those further in. In fact, the reverse is the case. The concentration of the excited states of atoms of a given element can be calculated as a function of height in the solar chromosphere, assuming the temperature to be that of the reversing layer (4830° K.) at its base where the Fraunhofer spectrum is formed. Eclipse observations of the chromospheric spectrum, however, have shown that many lines, especially those of hydrogen and helium, extend to several times the predicted heights. A higher temperature in the chromosphere is evidently required, and a value of the order of $30,000^{\circ}$ K. has been confirmed spectroscopically by Redman, who attributes the observed widths of the chromospheric hydrogen lines to thermal broadening.

Above the chromosphere is the corona. Its spectrum could not be more different. The emission lines, as studied by Lyot and Waldmeier and interpreted by Edlén, are those of highly ionized metals. Three independent experimental determinations agree upon a value for the kinetic temperature of the coronal material of the order of 10^6 deg. K. How then is this inverted temperature gradient maintained? Hoyle's reply is the bold, yet obvious, assertion that energy is being supplied to the top of the coronal atmosphere by the inflow of interstellar material under the action of gravitational attraction.

Accretion of interstellar material is no new idea; but the theory has never been worked out in all its implications and with such a wealth of quantitative detail as has been done by Mr. Hoyle and his collaborators. It is the main theme of the book, as set out in Chapters III and IV, but there is much more in addition. The author has a genius for sifting through the observational material and for selecting what are the most significant features in relation to his own special problems. Thus Chapters I and II are mainly concise summaries of the observational data concerning sunspots, the solar cycle, prominences, chromosphere and corona. In Chapter V he puts forward what promises to be a very fruitful theory of prominences, regarding them as cooler regions of the corona where the temperature is low enough for many of the normal chromospheric lines to become visible. There are new ideas on the sunspot magnetic cycle, and Giovanelli's work on solar flares is described and extended. Chapters V and VI deal with solar and terrestrial relationships and with the emission of radio waves from the sun.

In an appendix which discusses the origin of the sun's general magnetic field the theory is advanced that the sun's present magnetic field may have originated within the last million years or so during one of the temporary periods of very rapid accretion of interstellar matter. Indeed, it is one of the virtues, if we may call it such, of the accretion hypothesis that it allows a certain degree of flexibility, so to speak, in the ordering of solar affairs, and evidence for this is not wanting. As the sun wends its dusty way through interstellar space it may sweep up the interstellar particles from time to time at very different rates. During periods of high accretion rate solar activity will be on a grander scale than at present. Mr. Hoyle thinks that during one of these periods the earth may have received its present magnetization through streams of ions ejected from solar flares far more intense than those which now generate terrestrial magnetic storms.

Enough has been said to convey the impression that this is a book full of new and challenging ideas: many of them have been worked out in full detail, others are no more than daring speculations which may, or may not, survive. However that may be, it is certain that this little volume contains a sufficient amount of solid achievement to place solar physicists in Mr. Hoyle's debt for many years to come.

M. A. ELLISON.

Atmospheric Turbulence, by O. G. SUTTON. Pp. viii+107. Methuen's Monographs on Physical Subjects. First Edition. (London: Methuen, 1949). 6s.

Professor Sutton has been a leader during the past two decades in the development of the British school on atmospheric turbulence and he gives us here a well-written account of that development. To the reader a little outside the subject he provides an opportunity to see how far along the road of understanding some parts of the subject have progressed, and, what is clearly not the same thing, how far the phenomena dealt with have been made amenable to mathematical treatment.

This is a monograph and, fittingly enough, the author has followed a personal bent in dealing with his subject wherever possible by mathematical methods. But one should hasten to add that no physicist should find the matter difficult to read on that account. He further limits himself in the main to a study of the problems of the turbulent diffusion of momentum, heat and matter in the lowest 100 metres or so of the atmosphere; a quite understandable choice since it is in that layer that knowledge has in the main advanced in recent years. Nevertheless in making the choice the author claims to be dealing "with the most important region for turbulence" and "to have indicated to the reader those parts of the subject . . . significant for future development."

The first chapter is mainly devoted to a qualitative description of the features of turbulent flow in general, and to analogies between turbulence and the random motion of molecules. Then follows a concise description of the observations of turbulent flow in the shallow layer of atmosphere accessible to earthbound instruments and of the vertical gradients and diurnal variations of mean quantities (wind, temperature, etc.) affected by and effecting the turbulence. In four remaining chapters the author develops a mathematical treatment of turbulent diffusion, based in turn on empirical K (eddy viscosity) theory, on Prandtl mixing-length theory, and on the author's own empirical formulation of the Taylor correlation function, the function which measures the rate at which an eddying mass of air loses its identity during motion through its environment.

The older K -theory is rather quickly dismissed as sterile, but mixing-length and correlation-function theory are used to erect a seemingly substantial structure of equations for the transfer and concentrations of momentum, heat and matter from various sources. The structure has, however, little cement between the members or in the foundations. The cement one looks for is satisfying physical argument or idea but whether it will ever be applied to this particular edifice is questionable. The formulae can undoubtedly be useful in practical application but their range of validity is strictly within the compass of their test in the field. This is roughly the present state of the subject and to say so is no criticism of Professor Sutton's matter, though one might have expected that he would have stressed more strongly the quite empirical nature of the work. It would be little exaggeration to say that practically the whole of the theory presented in this monograph, Taylor's theorem of diffusion apart, is derivable from dimensional analysis and while he would be a poor physicist who spurned such aid, dimensional analysis is not enough to provide a very lively body of knowledge. The author indicates that ideas for a more or less fresh approach are beginning to emerge but, as he says, the ideas are almost still confined to the fluid-motion laboratory and can hardly yet face the atmosphere with its complicating density gradients and larger scales of motion.

Four-fifths of the earth's surface being water, one would have liked to find some reference in a treatment of terrestrial boundary-layer turbulence to the lowest 100 metres or so over the sea, for this boundary undoubtedly has special properties. Perhaps, when a second edition is called for, the author will enlarge his treatment in this sense.

The book is well printed, referenced and indexed, and will undoubtedly find a welcome as a most competent survey of a useful, if not very meaningful, phase of development in this entrancing but tantalizing subject.

P. A. S.

Practical Microscopy, by L. C. MARTIN and B. K. JOHNSON. Pp. 124. Blackie's "Technique" Series. 2nd Edition. (London and Glasgow: Blackie and Son, Ltd., 1949.) 6s. 6d.

Any book which sets out to describe the technique of modern microscopy in 124 small pages must of necessity be incomplete in some respects. Let it be said at once that "Martin and Johnson" is one of the best books of its kind. It explains the basic principles in a clear and simple manner, yet manages to deal with many small points which are not covered by much larger volumes. The practical microscopist will find of particular value the many numerical examples to illustrate common problems, and the chapter on optical artefacts is one which cannot be read too often. It is easy to criticize omissions in a book of this size. One cannot help expressing some disappointment that a rather more thorough revision has not been undertaken. True, a chapter on the electron microscope and a section on phase contrast illumination have been added, but apart from this the book has scarcely been altered since 1931. Thus it is still stated that "the use of the polarizing microscope in biological work does not appear to have received the attention it deserves", whereas the literature on the subject now runs into thousands of papers, and the instrument itself has been vastly improved. The basic principles of ultra-violet microscopy are well described, but modern developments, in which one of the authors has played an important part, are not referred to.

Despite these criticisms, which the reviewer hopes will be remedied in the many future editions which the book deserves, "Martin and Johnson" remains a book which should be available to every microscopist.

R. BARER.

Ionization Chambers and Counters: Experimental Techniques, by B. B. ROSSI and H. H. STAUB. Pp. xviii+243. 1st Edition. (New York, Toronto, London: McGraw-Hill Book Co. Inc., 1949.) 19s. 6d.

This volume is the second of the National Nuclear Energy Series (Los Alamos Project, Division V) and it has an obviously close association with the first—*Electronics in Nuclear Physics*—by E. L. Elmore and M. L. Sands. The first four chapters deal with the fundamental principles of ionization and the general properties of detectors based upon the ionization process. Many readers will find Chapter 1 of special interest since it collects much useful data on the drift velocity and agitation energy of electrons; a good deal of material due to the authors themselves appears here for the first time and it refers closely to counting techniques. Chapters 2 and 3 are concerned with the general design considerations affecting the performance of chambers and the latter deals in very thorough fashion with the relatively new problem of the measurement of variable ionization. In Chapter 4 the basic ideas of the gas multiplication process are rather briefly presented. A series of curves gives the values of multiplication observed in a variety of gases in typical operating conditions. The studies of pulse shape, end-effect of tubes and multiple-wire arrangements follow. While the allocation of space to the different sections seems somewhat strange in a textbook all of it is of interest to the specialist in this field.

The last five chapters consider in some detail the design of detectors for beta-, gamma- and x-radiations (Chapter 5), alpha-particles, neutrons and fission products. Chapter 5 is comparatively modest in its scope and rather unbalanced. On the other hand, the compact discussion of α -ray spectroscopy is very meritorious. The two long chapters devoted to the neutron form the core of this section of the book; the whole subject is treated with originality and the authors are to be congratulated on their thorough treatment of the many difficulties encountered in this field. Again, a considerable fraction of the material is original. Finally, Chapter 10 discusses fission detectors.

The book suffers in several directions due to its origin and it is particularly disappointing that there is no bibliography. It is a notable contribution to the very limited number of texts on a subject of major importance in atomic and nuclear physics. The number of engineering drawings is unusually generous. All closely interested in the subject will be grateful to the authors for this excellent book.

S. C. CURRAN.

Differential Equations, by HARRY W. REDDICK. Pp. x+288. 2nd Edition. (New York: John Wiley and Sons; London: Chapman and Hall, Ltd., 1949.) 24s.

This is a relatively elementary textbook, not, as its title implies, on differential equations generally, but on ordinary (as opposed to partial) equations. It could certainly be understood and used immediately after a first course in calculus, and might be useful to those who intend to be engineers or applied scientists, but not to those who intend to take their mathematical physics to any high standard. For the latter, the approach is much too lacking in rigour, a feature which is illustrated in the first chapter, where the difficulties associated with deciding how many arbitrary constants must appear in the solution of an equation of given order are glossed over. It is not a criticism of the book, but an indication of its scope, to say that existence theorems are lacking and that Sturm–Liouville systems are not mentioned, nor are their typical properties illustrated even in a simple particular case.

Within its restricted aims, however, the book is clear and accurate, and has the excellent feature of containing many good examples, with answers provided. It should serve well to teach the tricks of integrating those equations which have solutions expressible in closed form or in power series.

J. H. A.

Formulas and Theorems for the Special Functions of Mathematical Physics, by WILHELM MAGNUS and FRITZ OBERHETTINGER; translated by JOHN WERMER. Pp. 172. (New York: Chelsea Publishing Co., 1949.) No price.

The special functions of mathematical physics are those satisfying a certain number of differential equations familiar in different parts of theoretical physics, and include Bessel functions, spherical harmonics, hypergeometric functions and elliptic functions. As specialized cases, there are Laguerre functions, Hermite functions and many others.

In this book, which is intended as a reference work and not as a textbook, the main (and a good many rather recondite) properties of a great number of these functions are set down in convenient form. There cannot be many of the properties given for the cylindrical functions which are not to be found in Watson's treatise, but it is probable that they could be found rather more easily in this compendium than in the larger work.

The standard of mathematical rigour is high, and the gamma function, for example, is defined in a way which applies to all (real or imaginary) values of its argument.

The book gains a good deal in value by containing a chapter (not strictly justified by the title) on integral transformations and one on coordinate transformations which, rather unexpectedly, does not refer to group theory.

J. H. A.

Détermination d'un état plan des contraintes à l'aide d'un extensiomètre à résistance électrique à trois directions (rosette): Abaques Pratiques d'Emploi, par F. T. SALLES. Pp. 41. (Paris: Publications Scientifiques et Techniques du Ministère de l'Air, No. B.S.T. 112, 1949.) 350 fr.

The use of resistor strain gauges for the measurement of stresses and strains in structures, etc., has now become common practice. If the stress-system is two-dimensional, it is necessary to measure the strains in three directions at a given point in order to determine the strains (or stresses) completely; one method of doing this is to use a rosette gauge, consisting of three independent, intersecting resistor gauges, cemented to the structure at the point in question. The reduction of the experimental results is tedious and, in this report, tables and nomograms are provided to facilitate this work when the three gauges of the rosettes are inclined at 45° or at 120° to each other (rectangular or equilateral rosettes, respectively). The publication should be very useful to workers who use the rosette type of gauge.

R. M. D.

CONTENTS FOR SECTION A

PAGE

PAGE

Prof. W. FRANZ. On the Theory of Diffraction	925
Miss M. E. PILLOW. Intensity Distribution in Band-Systems of O_2 and O_2^+	940
Dr. S. P. SINHA. Ultra-Violet Bands of K_2	952
Mr. E. B. ANDREWS and Dr. R. F. BARROW. A New Ultra-Violet Band-system of GeBr	957
Mr. A. HEDGRAN, Dr. K. SIEGBAHN and Dr. N. SVARTHOLM. A Large β -Spectrometer with Two-Directional Focusing for Precise Measurements of Nuclear Radiation	960
Dr. D. K. BUTT. Experiments with a Magnetic Lens Spectrometer, employing Post-focusing Acceleration : the L-Auger Lines of Thorium-active Deposit	986
Mr. C. H. COLLIE, Dr. H. HALBAN and Dr. R. WILSON. Photoelectric Dissociation of the Deuteron	994
Prof. L. JÁNOSSY. On the Lateral Spread of Extensive Air Showers	1009
Mr. VACHASPATHI. On the Use of β -Formalism of the Meson Field for Nuclear Interactions	1015
Dr. KAI-CHIA CHENG. A New Method for Determining the Radial Distribution Function	1028
Mr. J. HAMILTON. Statistical Equilibrium and Radiation Damping	1036
Letters to the Editor :	
Mr. E. RABINOWICZ. A Photographic Method for determining Half-Lives	1040
Dr. C. G. B. GARRETT. An Order-Disorder Transition Curve in Cobalt Ammonium Sulphate below $0.1^\circ K$	1042
Mr. E. W. FULLER. The Half-Life of the Isomeric State in ^{109}Tm	1044
Dr. J. B. BIRKS. Scintillations from Naphthalene-Anthracene Crystals	1046
Reviews of Books	1046
Corrigendum	1050
Contents for Section B	1050
Abstracts for Section B	1051

ABSTRACTS FOR SECTION A

On the Theory of Diffraction, by W. FRANZ.

ABSTRACT. A new method is described for calculating the diffraction of an acoustical or electromagnetic wave by successive approximations. Kirchhoff's theory is a special case of the first step in the new method, which, unlike Kirchhoff's, is not restricted to black screens, and applies to long waves as well as short ones.

Intensity Distribution in Band-systems of O_2 and O_2^+ , by M. E. PILLOW.

ABSTRACT. The intensity distributions for three oxygen band-systems, the Schumann-Runge and Herzberg systems of O_2 and the second negative system of O_2^+ , are calculated, and compared where possible with experimental results. Some suggested identifications of bands observed in emission from the night sky are checked by comparison with the calculations.

Ultra-Violet Bands of K₂, by S. P. SINHA.

ABSTRACT. Bands of K₂ have been photographed in absorption in the region λ 3924– λ 3420 Å. on a Hilger E1 quartz spectrograph. The bands lie in two groups, one between λ 3924 and λ 3686 Å. and the other between λ 3640 and λ 3420 Å., and appear to belong to two different systems.

The bands in the first group have been analysed. They can be represented by the equation :

$$\nu = 26493.0 + 60.6u' - 0.15u'^2 - 92.64u'' + 0.354u''^2,$$

where $u = v + \frac{1}{2}$.

The upper state of this system is considered to dissociate into 4²S + 5²P atoms of potassium.

No analysis for the bands lying between λ 3640 and λ 3420 Å. has been possible.

A New Ultra-Violet Band-System of GeBr, by E. B. ANDREWS and R. F. BARROW.

ABSTRACT. The vibrational analysis of a new system of GeBr has been made. The constants derived in cm⁻¹ are :

	ν_0	ω_0	$x_0\omega_0$
² Δ or ² Π	27252	197.3	7.18
	27156	190.1	7.28
² Π _{3/2}	1150		
² Π _{1/2}	0	295.4	0.72

The upper state, probably ²Δ or ²Π, appears to converge rather rapidly to a limit at about 3.5 ev. above $v''=0$ of ²Π_{1/2}.

A Large β-Spectrometer with Two-Directional Focusing for Precise Measurements of Nuclear Radiation, by A. HEDGRAN, K. SIEGBAHN and N. SVARTHOLM.

ABSTRACT. In the present paper a large double focusing spectrometer is described, which is convenient for precise measurements of β- and γ-radiation. The magnetic field form is discussed and also the arrangement to measure the value of the field continuously with an accuracy appropriate to investigations of this kind. The difficulties due to electron straggling in the β-source or converter are emphasized, and this effect is for the moment the factor which prevents the full utilization of the high resolving power of the apparatus. A number of typical spectra are given to illustrate the applicability of the instrument to different cases.

Experiments with a Magnetic Lens Spectrometer, employing Post-focusing Acceleration : the L-Auger lines of Thorium-active Deposit, by D. K. BUTT.

ABSTRACT. A post-focusing electron accelerator has been developed to reduce the absorption of low-energy electrons by the Geiger counter window of a β-ray lens spectrometer. Its useful range is for β-rays having energies between 0 and 30 kev., and although, in general, for a portion of this range only partial transmission will be obtained, correction curves for the absorption can be accurately determined.

In order to test the spectrometer, an investigation of the L-Auger lines of thorium-active deposit has been made. An identification of the lines has been attempted.

Photoelectric Dissociation of the Deuteron, by C. H. COLLIE, H. HALBAN and R. WILSON.

ABSTRACT. The cross section for the photo-disintegration of the deuteron has been measured by detecting the photo-protons formed in a high pressure ionization chamber. The values obtained are $13.91 \pm 1 \times 10^{-28}$ cm² for 2.62 mev. γ-rays from radio-thorium and $15.6 \pm 1 \times 10^{-28}$ cm² for the 2.76 mev. γ-rays from radio-sodium.

On the Lateral Spread of Extensive Air Showers, by L. JÁNOSSY.

ABSTRACT. The lateral and angular spread of extensive showers is calculated in an approximation where ionization loss and incomplete screening are neglected (approximation A). The atmosphere is taken to be homogeneous. It is shown that the spread at the cascade maximum is equal to the average spread, but the spread increases considerably throughout the shower; the above results are in qualitative agreement with results given by Borsellino. The quantitative discrepancy between the results of Borsellino and other authors is commented on.

On the Use of β -Formalism of the Meson Field for Nuclear Interactions, by VACHASPATI.

ABSTRACT. Following a method due to Harish Chandra, a correspondence between the Proca and the β -formalisms of the meson field theory is developed. An important feature of the method is that one can at any stage of calculations pass over from one formalism to the other. No explicit representation of the β -matrices is required nor need one know the exact relationship between the elements of the wave function ψ and the Proca field quantities U_μ and $G_{\mu\nu}$. This introduces considerable simplification in the introduction of nuclear interaction as compared with earlier treatments where one has to choose a special explicit representation of the β_μ 's and identify the various elements of the matrix ψ with the components of U_μ and $G_{\mu\nu}$. The nuclear interaction is, in fact, here introduced very simply in an elegant way. The quantization is performed for the vector, scalar and pseudoscalar cases assuming in all cases both charge and dipole types of couplings. As an illustration of the method we calculate the scattering of positively charged mesons by protons. The spur calculations, so tedious in the earlier works, are now replaced by simple relations which are easily handled, and the results are obtained with about as much labour as is used in Proca's case with the additional advantage that both the longitudinal and the transverse parts of the meson field can here be treated together.

A New Method for Determining the Radial Distribution Function, by KAI-CHIA CHENG.

ABSTRACT. It is shown in this paper that Green's integral equation resulting from the Kirkwood approximation for the radial distribution function can be transformed into a differential equation of infinite order, to which the method of successive approximation can be applied without ambiguity. The solution at moderately high temperatures is discussed. The result is a correction to the usual Boltzmann distribution function; the radial distribution function oscillates with decreasing amplitudes at small distances and tends to unity at large distances. The validity of the approximations is discussed in two appendices.

Statistical Equilibrium and Radiation Damping, by J. HAMILTON.

ABSTRACT. The reason why the usual form of radiation damping theory does not satisfy the principle of detailed balancing is discussed. A form of damping theory satisfying statistical equilibrium is developed.

THE PROCEEDINGS OF THE PHYSICAL SOCIETY

Section B

VOL. 63, PART 10

1 October 1950

No. 370 B

The Influence of the Condenser on Microscopic Resolution

By H. H. HOPKINS AND P. M. BARHAM

Imperial College, London

MS. received 13th April 1950

ABSTRACT. An expression is found for the distribution of light in the image of two small apertures which are illuminated by a condenser of numerical aperture equal to s times that of the imaging objective. The resolving power of the system is written $\rho = K\lambda/(N.A.)$, and K is found as a function of s . It is shown that precisely the same result obtains for both critical and Köhler illumination.

§ 1. INTRODUCTION

THE theory of the resolving power of a telescope is generally based on a consideration of the image of a double star. Optically this is represented as the image of two closely spaced incoherent point sources of the same intensity. Each source produces a diffraction pattern in the focal plane of the telescope objective. The image of the double star is then obtained as the sum of the separate intensities in the two diffraction patterns. The Rayleigh criterion of resolution requires that the minimum of intensity midway between the geometrical images is of the order of 20% smaller than the intensity at the maxima of the summed intensities. Conventionally the criterion is regarded as being satisfied when the geometrical image of one source falls on the first dark ring of the diffraction image of the other.

In the case of a slit shaped aperture the intensity distribution in the image of a point source has the form $I = \{\sin \pi x/x\}^2$. This gives a decrease in intensity of 18.9% at the point midway between the geometric images. If a circular aperture is employed, the intensity distribution has the form $I = \{2J_1(x)/x\}^2$. The decrease in intensity is then found to be 26.4%. If α is the angular semi-aperture, and N a refractive index, the two sources will be 'resolved' when their geometrical images are separated by a distance ρ , where

Slit shaped aperture: $\rho = 0.50\lambda/N \sin \alpha$ (1)

Circular aperture: $\rho = 0.61\lambda/N \sin \alpha$ (2)

It is significant that the smaller resolvable separation (1) is associated with a smaller drop in intensity.

Disturbances from two distinct point sources are incoherent. In the case of a telescope imaging double stars, it therefore suffices to add the intensities of the separate diffraction patterns. This will not generally be true for the images

of, say, two illuminated pinholes in an opaque screen, for there some degree of phase correlation will obtain. In general, therefore, it will be necessary to take phase relationships into account when adding the disturbances in the image plane of a microscope.

Martin (1931) considered the problem of the image of two small illuminated apertures. His treatment was only approximate, but showed very clearly the physical processes involved. The case of critical illumination was considered. Firstly assuming a point source for the illuminant, and then extending the treatment qualitatively to consider an extended source imaged on the object by means of a perfectly corrected condenser of aperture equal to that of the objective, it was concluded that "the ultimate effect is likely to be similar to that characteristic of two independent elementary sources", the two apertures being at a separation given by (2). In the present paper it is shown that this is, in fact, exactly true for both critical and Köhler illumination. This accords with a result predicted by Zernike (1938).

In our case an expression is obtained for the disturbance in the final image plane associated with the image of two pinholes illuminated by a condenser of any given aperture when used in conjunction with a large source of uniform intensity. Circular apertures are assumed for both condenser and objective. The resolvable separation ρ is then written

$$\rho = K\lambda / N \sin \alpha \quad \dots\dots (3)$$

instead of (2). K is found as a function of s , the ratio (N.A. condenser)/(N.A. objective). N.A. denotes, as usual, the numerical aperture equal to $N \sin \alpha$.

Recently a criterion of resolution due to Sparrow (1916) has been employed (Ramsay, Cleveland and Koppius 1941, Françon 1950). If I is the intensity at the point midway between the two geometric images, and x is a coordinate measured along the line joining them, the images are said to be resolved when $\partial^2 I / \partial x^2 = 0$. It has been claimed that this gives a true 'limit' of resolution, since the intensity curve is then 'flat'. However, the criterion is not unique, since the equation $\partial^2 I / \partial x^2 = 0$ will frequently have an infinite number of roots. Moreover the 'resolution' indicated by this criterion is meaningless in some cases, notably in finding the limit of detection of the microscope. The diameter of that smallest particle that can be 'seen' is found to be zero. With increasing contrast sensitivity of the screen on which the image is received, the size of the smallest detectable particle does decrease indefinitely. In a sense, therefore, the result obtained using the criterion $\partial^2 I / \partial x^2 = 0$ is valid. Nevertheless it is of little practical value.

Because of this we have adopted a Rayleigh criterion of resolution which for circular apertures requires the decrease in intensity to be 26.4% of the maximum intensity. The analysis is based on a consideration of critical illumination. It is then shown that the same result obtains for Köhler illumination.

§ 2. CRITICAL ILLUMINATION WITH AN EXTENDED SOURCE

We have first to define an optical unit of length. Suppose a length ρ in the object plane has an image of length ρ' . Then the magnification is given by $m = \rho' / \rho$. If α, α' are the angles which a ray makes with the axis in the object and image spaces respectively, and these latter have refractive indices equal to

N and N' , the magnification is also given by $m = N \sin \alpha / N' \sin \alpha'$. Equation of these two expressions shows that the one coordinate length

$$z = \frac{2\pi}{\lambda} \rho N \sin \alpha = \frac{2\pi}{\lambda} \rho' N' \sin \alpha' \quad \dots\dots(4)$$

can be used to define equally ρ or ρ' . This is convenient since it is the variable (4) which appears in the functions expressing the amplitude in the diffraction distributions.

Consider now an opaque plane containing two small apertures, P_1, P_2 . The sizes of P_1 and P_2 will be assumed to be equal and both small compared with the Rayleigh resolution of the objective O , by means of which the two points are imaged at P_1', P_2' . Let α, α' in (4) refer to the objective O , and let $\rho_0 = P_1 P_2$, $\rho_0' = P_1' P_2'$. Then

$$z_0 = \frac{2\pi}{\lambda} \rho_0 (N.A.)_O \quad \dots\dots(5)$$

is the optical separation of P_1 and P_2 . $(N.A.)_O = N \sin \alpha$ is the numerical aperture of the objective.

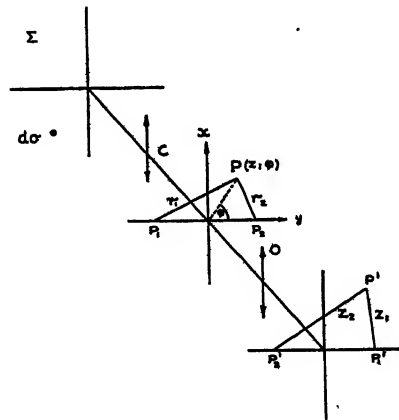


Figure 1.

If an element $d\sigma$ of the source Σ is imaged at $P(x, \phi)$ by the condenser C , the amplitude associated with the image of $d\sigma$ will be that in an Airy disc with centre at P . Let $\rho_1 = PP_1$, $\rho_2 = PP_2$. Then the amplitude at P_1 due to $d\sigma$ may be written

$$u_1 = \frac{J_1 \left\{ \frac{2\pi}{\lambda} \rho_1 (N.A.)_C \right\}}{\frac{2\pi}{\lambda} \rho_1 (N.A.)_C} = \frac{J_1(sr_1)}{sr_1} \quad \dots\dots(6)$$

where $(N.A.)_C$ is the numerical aperture of the condenser, and $s = (N.A.)_C / (N.A.)_O$. The variable r_1 is defined by $r_1 = (2\pi/\lambda) \rho_1 (N.A.)_O$. It is the optical measure of ρ_1 for the objective. The amplitude produced by $d\sigma$ at P_2 is given by

$$u_2 = \frac{J_1(sr_2)}{sr_2} \quad \dots\dots(7)$$

where r_2 is similarly defined for ρ_2 . The condenser is assumed to be free from aberration and perfectly focused.

The diffraction images of the points P_1, P_2 will be Airy discs centred on P_1', P_2' respectively. The amplitude produced at P' by disturbances originating at P_1, P_2 will therefore be given by

$$u = u_1 \frac{J_1(z_1)}{z_1} + u_2 \frac{J_1(z_2)}{z_2},$$

where z_1, z_2 are the optical distances $P_1'P', P_2'P'$. The disturbances u_1, u_2 are coherent, since they derive from the same element $d\sigma$ of the source. To find the intensity at P' due to the whole source requires the integration of u^2 over the domain Σ . That is

$$I(z_1, z_2) = \int_{\Sigma} \left\{ u_1 \frac{J_1(z_1)}{z_1} + u_2 \frac{J_1(z_2)}{z_2} \right\}^2 d\sigma. \quad \dots\dots (8)$$

Expanding the expression in brackets in (8),

$$I(z_1, z_2) = \left\{ \int_{\Sigma} u_1^2 d\sigma \right\} \frac{J_1^2(z_1)}{z_1^2} + \left\{ \int_{\Sigma} u_2^2 d\sigma \right\} \frac{J_1^2(z_2)}{z_2^2} + 2 \left\{ \int_{\Sigma} u_1 u_2 d\sigma \right\} \frac{J_1(z_1) J_1(z_2)}{z_1 z_2}. \quad \dots\dots (9)$$

Using the expressions (6) and (7) for u_1 and u_2 the integrals in (9) may be evaluated. To do this the domain Σ is assumed to be of infinite extent. Physically this requires that the geometrical image of the source shall be large compared with the distance P_1P_2 .

Let P_1, P_2 be at the points $(-y, 0), (+y, 0)$ where $2y = z_0$. (z, ϕ) are the coordinates of P . Then

$$r_1^2 = y^2 + z^2 + 2yz \cos \phi, \quad r_2^2 = y^2 + z^2 - 2yz \cos \phi$$

and the integral in the first term of (9) is

$$\int_{\Sigma} u_1^2 d\sigma = \int_0^{\infty} z dz \int_0^{2\pi} \left\{ \frac{J_1\{s(y^2 + z^2 + 2yz \cos \phi)^{1/2}\}}{s(y^2 + z^2 + 2yz \cos \phi)^{1/2}} \right\}^2 d\phi.$$

The domain of integration has been formally extended to the infinite plane. The integral remains unchanged if the origin is displaced to the point $(-y, 0)$. It is then given by

$$\int_{\Sigma} u_1^2 d\sigma = \int_0^{\infty} z \left\{ \frac{J_1(sz)}{sz} \right\}^2 dz \int_0^{2\pi} d\phi = \frac{\pi}{s^2} \quad \dots\dots (10)$$

since
$$\int_0^{\infty} \frac{J_1^2(t)}{t} dt = \frac{1}{2}.$$

By a similar procedure the integral in the second term of (9) is found to be

$$\int_{\Sigma} u_2^2 d\sigma = \pi/s^2. \quad \dots\dots (11)$$

There remains the integral in the third term.

This third integral is

$$\int_{\Sigma} u_1 u_2 d\sigma = \int_0^{\infty} z dz \int_0^{2\pi} \frac{J_1\{s(y^2 + z^2 + 2yz \cos \phi)^{1/2}\}}{s(y^2 + z^2 + 2yz \cos \phi)^{1/2}} \frac{J_1\{s(y^2 + z^2 - 2yz \cos \phi)^{1/2}\}}{s(y^2 + z^2 - 2yz \cos \phi)^{1/2}} d\phi \quad d\phi$$

which may be written

$$\int_{\Sigma} u_1 u_2 d\sigma = \int_0^{\infty} z dz \int_0^{2\pi} \frac{J_1(sz)}{sz} \cdot \frac{J_1\{s(z_0^2 + z^2 - 2z_0 z \cos \phi)^{1/2}\}}{s(z_0^2 + z^2 - 2z_0 z \cos \phi)^{1/2}} d\phi \quad \dots\dots(12)$$

if the origin is displaced to the point $(-y, 0)$, and $2y$ is replaced by z_0 .

The integrand in (12) may be expanded with the aid of Neumann's addition theorem for the Bessel functions (Watson 1942). Thus, consider the differentiation

$$\frac{\partial}{\partial \phi} \{J_0(R)\} = -\frac{J_1(R)}{R} ab \sin \phi,$$

where $R = \{a^2 + b^2 - 2ab \cos \phi\}^{1/2}$. Neumann's theorem has the form

$$J_0(R) = J_0(a)J_0(b) + 2 \sum_{p=1}^{\infty} \cos p\phi J_p(a)J_p(b).$$

Differentiating this last with respect to ϕ , and substituting for $(\partial/\partial \phi)\{J_0(R)\}$, there is found the result

$$\frac{J_1(R)}{R} b = \frac{2}{a} \sum_{p=1}^{\infty} \frac{p \sin p\phi}{\sin \phi} J_p(a)J_p(b).$$

If $a = sz_0$, $b = sz$, substitution of this expression in (12) gives

$$\int_{\Sigma} u_1 u_2 d\sigma = \frac{2}{s^2 z_0} \int_0^{\infty} \frac{J_1(sz)}{z} dz \int_0^{2\pi} \sum_{p=1}^{\infty} \frac{p \sin p\phi}{\sin \phi} J_p(sz_0)J_p(sz) d\phi.$$

The integral in ϕ has the value 2π when p is odd, and is zero when p is even.

That is

$$\int_{\Sigma} u_1 u_2 d\sigma = \frac{4\pi}{s^2} \sum_{p=1,3,\dots}^{\infty} p \frac{J_p(sz_0)}{sz_0} \int_0^{\infty} \frac{J_1(sz)J_p(sz)}{z} dz.$$

The remaining integral in z is one of Lommel's integrals (Watson 1942). It has the value zero for odd integral values of p , except $p=1$. In this case it is equal to $\frac{1}{2}$. Hence

$$\int_{\Sigma} u_1 u_2 d\sigma = \frac{\pi}{s^2} \frac{2J_1(sz_0)}{sz_0}, \quad \dots\dots(13)$$

the remaining terms in the p summation being zero.

Using (10), (11) and (13) in (9), the intensity at the point P' is found to be

$$I(z_1, z_2) = \frac{J_1^2(z_1)}{z_1^2} + \frac{J_1^2(z_2)}{z_2^2} + 2 \frac{J_1(sz_0)}{sz_0} \frac{J_1(z_1)}{z_1} \frac{J_1(z_2)}{z_2} \quad \dots\dots(14)$$

from which the common factor π/s^2 has been omitted.

Some interesting conclusions follow at once from (14). If sz_0 is a root of $J_1(sz_0)=0$, other than $sz_0=0$, the product term is absent. The intensity $I(z_1, z_2)$ is then expressed by the sum of the separate intensities associated with P_1 and P_2 . P_1, P_2 are therefore incoherently illuminated. In particular, this will be so, if $s=1$, when z_0 is a non-zero root of $J_1(z_0)=0$: that is, when the points P_1, P_2 are separated by a distance equal to the radius of any dark ring of the Airy pattern associated with the objective aperture. When P_1, P_2 are on the limit of resolution (2), these points will therefore be incoherently illuminated if the numerical aperture of the condenser is equal to that of the objective. Under these conditions the resolution of two illuminated small apertures is given by the expression (2).

If the aperture of the condenser is made limitingly small, (14) becomes

$$I(z_1, z_2) = \left\{ \frac{J_1(z_1)}{z_1} + \frac{J_1(z_2)}{z_2} \right\}^2 \quad \dots\dots(15)$$

since $\lim_{s \rightarrow 0} \{2J_1(sz_0)\}/(sz_0) = 1$. The points P_1, P_2 are then coherently illuminated no matter what their separation, z_0 . The significance of this in practice is best seen by noting that $2J_1(sz_0)/sz_0 = 0.880$ when $sz_0 = 1$. If $s = \frac{1}{10}$, points in a circle of radius $z_0 = 10$ will be very nearly coherently illuminated. The diameter of this coherently illuminated area ($2z_0 = 20$) is approximately five times the limit resolution ($z_0 = 3.83$) of the objective for two incoherent point sources.

To investigate the intensity along the line $P_1'P_2'$, a coordinate x measured from the mid-point of $P_1'P_2'$ may be used. The intensity at $P'(x)$ is then given by

$$I(x) = \frac{J_1^2(x+y)}{(x+y)^2} + \frac{J_1^2(x-y)}{(x-y)^2} + 2 \frac{2J_1(sz_0)}{sz_0} \frac{J_1(x+y)}{x+y} \frac{J_1(x-y)}{x-y}, \quad \dots\dots(16)$$

where $y = \frac{1}{2}z_0$ as before.

It is a simple matter, using (16), to find the separations of P_1P_2 which give a decrease in intensity of 26.4% at O, when s has different values. In this manner the value of K has been found for the range $s = 0.0 - 2.0$. The curve of Figure 2,

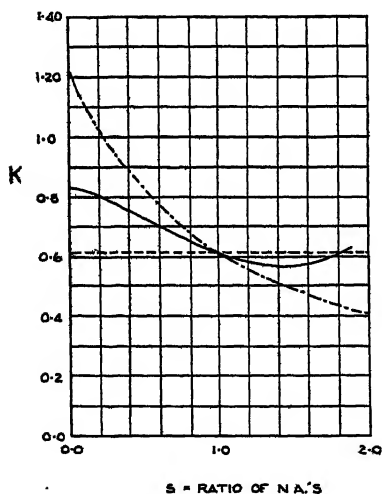


Figure 2.

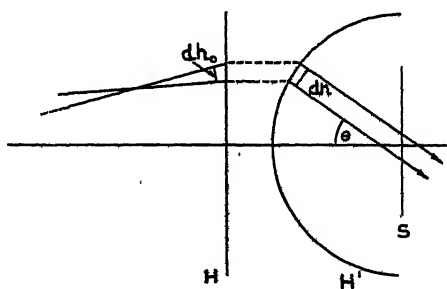


Figure 4.

shows the variation of K with s . If the influence of the aperture of the condenser on resolution is ignored, $K = 0.61$ for all values of s . This is the broken straight line in Figure 2. The broken curved line shows K as a function of s on the basis of the rule attributed to Abbé. According to this the effective aperture is the mean of the apertures of the condenser and the objective. It leads to an absurd result if the numerical aperture of the objective tends to zero. On the other hand it gives a rough approximation to the present result if $0.5 < s < 1.0$. When one remembers the other factors (such as scattered light, contrast in the object) which influence the resolution of the microscope, it is not surprising that the Abbé rule has been acceptable in practice.

§ 3. KÖHLER ILLUMINATION

In critical illumination coherent light from an element of the illuminant is found in the object plane as an Airy distribution of amplitude. Thus a single element of the illuminant illuminates any two points in the object co-phasally, but with different amplitudes. In Köhler illumination coherent light from an element of the illuminant illuminates two points in the object with light of the same amplitude but having a phase difference. The arrangement is shown diagrammatically in Figure 3.

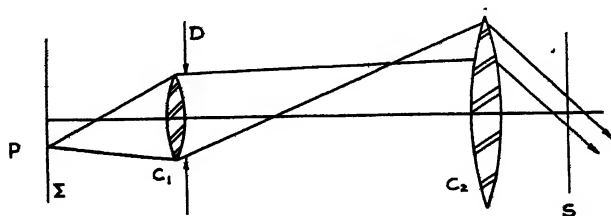


Figure 3.

The source Σ is imaged by a condenser C_1 in the focal plane of a second condenser C_2 , which is further arranged to image a diaphragm D in the plane of the object S . Light from an element P of Σ therefore results in a plane wave falling obliquely on the axial region of the object plane S . The object is illuminated by a set of plane waves whose normals lie within the cone subtended by C_2 at S . Let this cone be of semi-angle β , and write $N \sin \beta = s(N.A.)_0$, $(N.A.)_0$ being the numerical aperture of the objective, and N the refractive index of the space between C_2 and S .

Consider two points in the object S , such as P_1, P_2 in Figure 1. The plane wave falling on S whose normal makes an angle θ with the axis in the azimuth ϕ will have a phase retardation $kN \sin \theta \cdot \rho \cos \phi$ at P_2 relative to the phase at P_1 . The disturbances at P_1 and P_2 are thus given by

$$\left. \begin{aligned} u_1 &= A, \\ u_2 &= A \exp(-ikN \sin \theta \cdot \rho \cos \phi), \end{aligned} \right\} \dots (17)$$

where A is the real amplitude at these two points.

The condensers are assumed to be free from aberration. They therefore satisfy the sine condition, and in consequence the rear unit surface of C_2 will be a sphere with its centre at the axial point of S . If A_0 is the amplitude of the incident wave at the unit surface H (Figure 4) of C_2 , and A is the amplitude of the emergent wave at the other unit surface H' , then the conservation of energy requires that $A_0^2 dh_0 = A^2 dh$ where dh_0, dh are elements of area. Now $dh = dh_0 / \cos \theta$, hence

$$A = A_0 \cos^{1/2} \theta \dots (18)$$

and the expressions (17) become

$$\left. \begin{aligned} u_1 &= A_0 \cos^{1/2} \theta, \\ u_2 &= A_0 \cos^{1/2} \theta \exp(-ikN \sin \theta \cdot \rho \cos \phi). \end{aligned} \right\} \dots (19)$$

Let P_1', P_2' be the images of P_1, P_2 (as in Figure 1) and let P' be a point (z_1, z_2) . Then the intensity at P' is given by

$$I(z_1, z_2) = \iint \left| u_1 \frac{J_1(z_1)}{z_1} + u_2 \frac{J_1(z_2)}{z_2} \right|^2 d\Omega \quad \dots\dots (20)$$

as in (8), but here u_1, u_2 have the values given in equation (19). $d\Omega$ is an element of solid angle equal to $\sin \theta d\theta d\phi$, and the limits of integration are $0-2\pi$ for ϕ and $0-\beta$ for θ . Further, since u_2 is complex, it is the squared modulus which appears in the integrand of (20). Hence corresponding to (9), (20) gives

$$\begin{aligned} I(z_1, z_2) = & \left\{ \int_0^\beta \int_0^{2\pi} \sin \theta \cos \theta d\theta d\phi \right\} \left\{ \frac{J_1^2(z_1)}{z_1^2} + \frac{J_1^2(z_2)}{z_2^2} \right\} \\ & + 2 \frac{J_1(z_1)}{z_1} \frac{J_1(z_2)}{z_2} \int_0^\beta \int_0^{2\pi} \cos(kN \sin \theta \cdot \rho \cos \phi) \cdot \sin \theta \cos \theta d\theta d\phi \\ & \dots\dots (21) \end{aligned}$$

a constant factor A_0^2 being omitted.

The integral in the first term of (21) has the value $\pi \sin^2 \beta$. The integral in the second term is equal to

$$2\pi \int_0^\beta \sin \theta \cos \theta J_0(kN\rho \sin \theta) d\theta;$$

writing $x = kN\rho \sin \theta$, and $sz_0 = kN\rho \sin \beta$, this integral becomes

$$\frac{2\pi \sin^2 \beta}{s^2 z_0^2} \int_0^{sz_0} x J_0(x) dx = \pi \sin^2 \beta \frac{2J_1(sz_0)}{sz_0}.$$

Omitting the factor $\pi \sin^2 \beta$, (21) is now

$$I(z_1, z_2) = \frac{J_1^2(z_1)}{z_1^2} + \frac{J_1^2(z_2)}{z_2^2} + 2 \frac{2J_1(sz_0)}{sz_0} \frac{J_1(z_1)}{z_1} \frac{J_1(z_2)}{z_2} \quad \dots\dots (22)$$

which is identical with (14). Hence integrating the phases in Köhler illumination leads to precisely the same result as integrating the amplitudes in critical illumination.

This result is of considerable interest in practical microscopy. It shows that, given a broad source of uniform intensity, the appearance of the image will be independent of whether Köhler or critical illumination is employed. The Köhler system has, however, two distinct advantages in other respects. There is no difficulty in obtaining uniform illumination of the object, and the area of the object in which light falls may be controlled by the diaphragm D. By this means scattered light can be reduced with considerable improvement in the contrast of the image.

It is unfortunate that critical and Köhler illumination are designated in French by the terms 'éclairage incohérente' and 'éclairage cohérente' respectively. The identity of the expressions (14) and (22) shows that the coherence between points in the object plane in the two cases must be exactly the same when the condensers have the same numerical aperture.

REFERENCES

- FRANCON, M., 1950, *Le contraste de phase* (Paris: Édit. Rev. Opt. Théor. Instrum.).
 MARTIN, L. C., 1931, *Proc. Phys. Soc.*, **43**, 186.
 RAMSAY, B. P., CLEVELAND, E. L., and KOPPIUS, O. T., 1941, *J. Opt. Soc. Amer.*, **31**, 26.
 SPARROW, C. M., 1916, *Astrophys. J.*, **44**, 76.
 WATSON, G. W., 1942, *Bessel Functions* (Cambridge: University Press), p. 134.
 ZERNIKE, F., 1938, *Physica*, **8**, 785.

Reflectivity of Thin Silver Films and their Use in Interferometry

By H. KUHN AND B. A. WILSON *

Clarendon Laboratory, Oxford

MS. received 31st March 1950

ABSTRACT. A direct photoelectric method allowed the reflectivity R and the transmissivity T of partially reflecting films to be measured to an accuracy of 0.3% of the incident light intensity. For a large number of silver films, deposited on glass by evaporation *in vacuo*, the values of R and T were measured for visible light of five different wavelengths, in the range of high reflectivities ($R > 0.7$). The change of R as a function of time was also studied. With only moderate control of the conditions of surface cleaning and evaporation, the values of R were nearly always found to lie close to a 'standard' curve giving R as a function of T , the wavelength, and the age of the film. Possible causes of scatter from this standard curve were investigated. T was found not to obey an exponential law as a function of the thickness.

On the basis of these results, resolving power and brightness of etalon fringes were plotted as functions of T . The conclusion is reached that, in most practical cases, the resolving power of etalons for visible light is limited by imperfections of the surfaces rather than by the reflectivity of the silver films, if their thickness is properly chosen.

§ 1. INTRODUCTION

IN instruments based on the interference of multiple beams, transparent silver films of high reflectivity are widely used. The Fabry-Perot etalon and interference filters are among the most important examples. The resolving power of these instruments is essentially proportional to $1/(1-R)$, where R is defined as the ratio of the intensities of reflected and incident light. In increasing the value of R by increasing the thickness of the film, a practical limit is set by intensity requirements, since the transmission factor T decreases rapidly. Beyond a certain limit, the resolving power can only be increased at considerable sacrifice of intensity. The accurate knowledge of the value of R which can be obtained for a given value of T at a given wavelength is therefore of practical value in interferometry.

Though a number of observers have carried out direct measurements of reflectivities of silver films, our knowledge of the values of R and T in the range of high reflectivity ($R > 0.75$) is far from adequate.

The discrepancies between results of different observers and the scatter of the values obtained by any one observer are very large, and it is unknown how far these differences are caused by errors of the measurements, which are generally as great as 1% and often more.

Measurements of widths of interference fringes have given lower limits of reflectivities but, especially for high values of R , the results are considerably affected by imperfections of the plate surfaces.

In order to study not only the values of the reflectivity as a function of transmission and wavelength, but also the scatter of the values of R for films deposited under similar conditions and the change of R with ageing, it was necessary to make the errors of measurement considerably smaller than these variations. This

* Rhodes Scholar.

aim was achieved in the experiments described below. They were mainly intended to provide an answer to the following questions : (i) What value of R can be obtained for any given value of T between 0 and 0.2 and any given wavelength of visible light, if a good standard technique of surface cleaning and evaporation is used, and what is the amount of scatter about this 'standard' value of R ? (ii) How does the value of R change in the course of time, if the film is kept under the same conditions as in ordinary spectroscopic use ?

Regarding the influence of the numerous factors of surface conditions, vacuum and speed of evaporation, some useful information was obtained, but a comprehensive study of this question was not attempted.

§ 2. THE OPTICAL METHOD

The principle of the method employed is simple and well known, but the accuracy of the results depends on some details which are not quite obvious and will therefore be described (Figure 1).

The plate P , carrying the film whose reflectivity is to be measured, is mounted on a spectrometer table, with the axis of rotation passing through the reflecting surface. The metal plate holding P has a circular aperture Q of diameter 1.1 cm. The light source S is in a fixed position, 200 cm. from P . A photo-multiplier M is

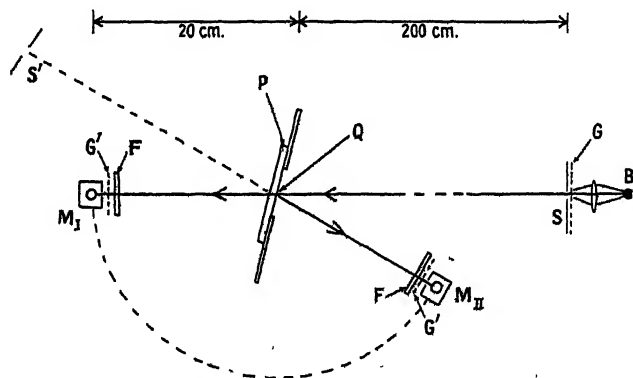


Figure 1. The optical system.

mounted on the movable arm of the spectrometer. In position M_I , the photo-multiplier receives light passing from S through Q directly when P is removed, or after passage through the plate P when this is inserted. In position M_{II} , it receives light reflected from P at a small angle of incidence (about 15°). This light can be regarded as coming from the virtual image S' of the source. The distance of S' from M_{II} is the same as that of S from M_I . The intensities observed in the three experiments are therefore in the ratio $1 : T : R$, provided the following conditions are fulfilled: (a) The light source and the aperture of the photocell must be so small that any ray passing from one to the other intersects the plane of P well inside the stop Q . (b) The light source must emit uniformly in all directions within the solid angle subtended by the aperture of the receiving system, and the latter must have an isotropic response. These conditions are especially important since S' is the mirror image of S , so that lack of uniformity in the angular distribution of the emitted light will have different effects in the two positions of M . (c) Light scattered by the mount of the reflecting plate, by the reflecting surface

itself, and by any other parts of the apparatus must be negligible. (d) The reflecting plate P must be sufficiently plane. (e) All geometrical adjustments have to be sufficiently accurate: the plane of P must pass through the axis of rotation, and the plane defined by S, M_I and M_{II} must be normal to the axis of rotation, and the lines SM_I and SM_{II} must pass approximately through the centre of Q.

In order to fulfil conditions (a) and (b), light from a 24-watt bulb B, connected to a large battery, was focused on a ground glass G which was fitted with a circular stop of diameter 5 mm. A similar ground glass G' and stop were fixed in front of the photo-multiplier.

The surface of the brass plate holding P was covered with black velvet paper. When M was set only slightly off the angle of reflection, the intensity fell to zero, showing practical absence of scattered light including light scattered from P itself. The intensity measured as a function of small horizontal or vertical displacements of M from the reflection position showed a symmetrical curve with a flat top whenever all adjustments had been carefully made.

The photo-multiplier, type R.C.A. I.P.21, was connected directly to a galvanometer. By varying the distance of the light source from the photocell and applying the inverse square law, it was found that the galvanometer deflection was not strictly proportional to the intensity, and all readings had to be corrected accordingly.

The effective wavelength range in each experiment was a function of the colour filter used, of the spectral intensity distribution of the source and of the response curve of the photosensitive layer. The latter has a maximum at about 4100 Å. and the sensitivity decreases rapidly in the near infra-red. From spectroscopic examination of the filters, the effective wavelength was estimated.

The Table gives a list of the filters used, the estimated limits of the effective wavelength bands, and the approximate centre of gravity of each band.

Ilford red	6400–7200 Å.	6800 Å.
Ilford green	5100–5400 Å.	5200 Å.
Ilford blue-green	4700–5200 Å.	4900 Å.
Wratten No. 48	4400–5000 Å.	4700 Å.
Wratten No. 35	3900–4500 Å.	4200 Å.

The violet, blue and blue-green filters were used in conjunction with a cell containing copper sulphate solution.

The colour filters F (Figure 1) were placed very close to the small aperture in front of the photo-multiplier. When they were placed further away, the small local variations of density of the filters were found to cause errors up to 0.5%.

When a measurement was repeated, after removal and re-setting of P and readjustment of the instrument, the value of R was generally reproduced to within 0.3%, and disagreements greater than 0.5% were practically never found. The estimated limits of error of a single measurement of R was thus about ± 0.003 . With the precautions described, all possible systematic errors should have been practically negligible, and the stated limits of error are believed to represent the absolute accuracy of the values of R . The accuracy of measurement of T is a little greater, so that the value of $R + T$ for each film should rarely be in error by more than ± 0.005 .

In the method described, the value of R was measured at an angle of incidence of 15° , whereas in the Fabry–Perot interferometer the angle of incidence is practically zero. For opaque metal films, the change of R with the angle of incidence can be

calculated from the values of the optical constants and is found to be of opposite sign for light polarized in the plane of incidence and at right angles to it. For the average of the two planes of polarization, the difference between the values of R for 0° and 15° is quite negligible (of the order of 0.001). For transparent films, the theoretical formulae are complicated, but for the comparatively low transmission used in our experiments, the dependence of R on the angle cannot differ much from that for thick films. This was indirectly confirmed by some measurements with plane-polarized light; the differences of the values of R for the different planes of polarization were small and close to the values predicted by the theory for thick films.

The error made in regarding the values of R measured at an angle of 15° as the reflectivity for normal incidence is below 0.002 for dense films, but it may arise to about 0.01 for the lowest reflectivities measured.

§ 3. THE PREPARATION OF THE FILMS

The evaporation chamber consisted of a large brass tank evacuated by means of a mercury diffusion pump of a rated speed of 25 litre/sec., but the liquid air trap and the duct to the chamber may have reduced the speed to about half this value. The pressure in the tank was measured by means of a Penning gauge. It was between 3 and 5×10^{-5} mm. Hg during the evaporation. These values refer to calibration with air.

Two tungsten filaments of thickness 0.7 mm., each carrying two beads of silver, could be heated independently. A few turns of thin platinum wire were wound round the V-shaped parts of the filaments, according to common practice, in order to cause the silver to adhere to the filaments. When a fairly good vacuum had been reached, air was admitted to a pressure of a few millimetres of mercury, and a discharge from a spark coil was passed through the tank for 5 to 10 minutes before the pumping was resumed. This treatment was repeated once or twice. The actual evaporation lasted about 5 minutes.

In each evaporation, three sets of two glass plates were used, at distances of 18, 20, and 22 cm. respectively from the filaments. This method allowed the comparison of films deposited under exactly the same conditions of evaporation.

The glass plates were of good optical quality, accurately plane and parallel, and of size 4.1 cm. \times 3.2 cm. Some of the plates were polished lightly with rouge before they were taken into use, so that the surfaces were similar to those of new plates, but this process was not repeated after the plates had once been covered with silver. The plates were washed in nitric acid, then rinsed with tap water and dried with cotton wool of good commercial quality. The surface was considered satisfactory when breathing produced a uniform film.

For the ageing tests, the plates were kept in a glass container, protected from dust by a loosely fitting cover. The container was left in a Laboratory room which was free from tobacco smoke, flames and chemical fumes.

§ 4. THE MEASURED VALUES OF R AND T

In the final set of experiments, the results of which are reported here, 18 glass plates were coated with silver in the way described and each plate was measured four times, first immediately after removal from the evaporation tank, then after intervals of 3, 8, and 22 days from the time of deposition. Each time the values of T and R were measured for five different wavelengths. Six plates

were left to be re-measured after 68 days, while the other 12 plates were cleaned, re-silvered and measured again when new and after 22 days.

For the measurements of these 30 films immediately after evaporation and after 22 days, the values of $R + T$ are plotted in Figure 2. Within the range of T

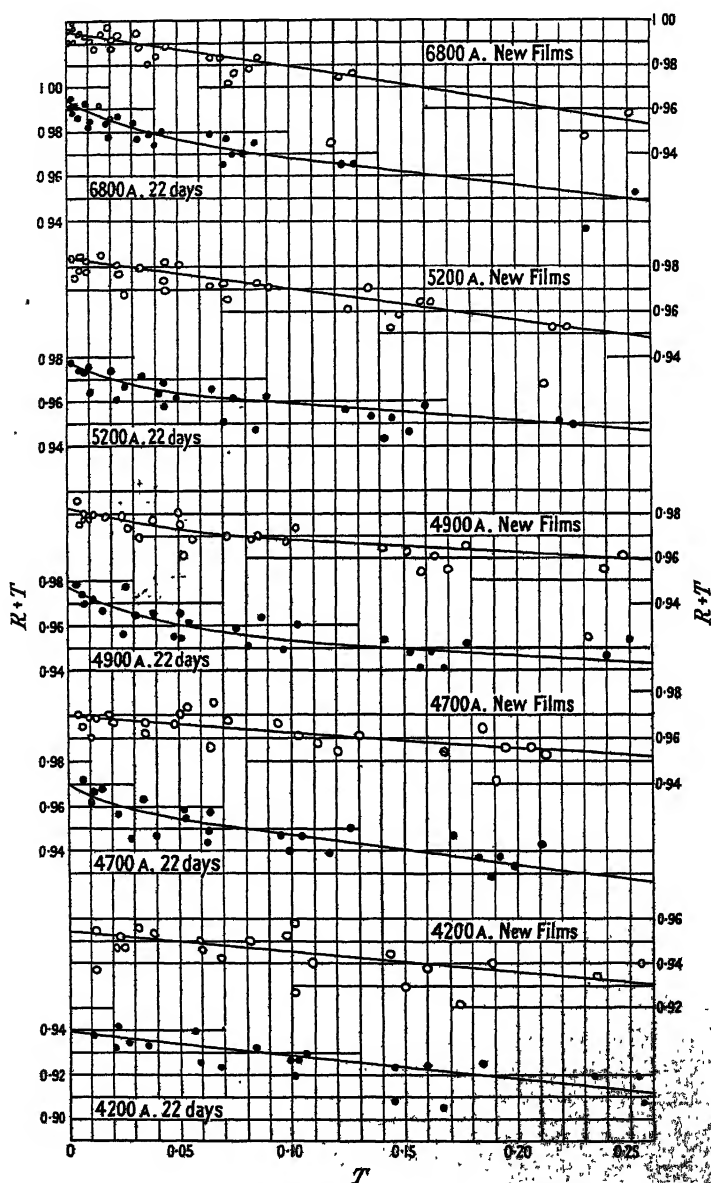


Figure 2. Measured values of $R+T$ as function of T .

plotted, these graphs contain all the measurements, without any selection. Only one film gave values so much below the curves (by about 5%) that they were not plotted for technical reasons, while for another film, for the same reasons, only the measurement in the red for the new film was included.

Since T changes with the wavelength, the same films do not appear at the same values of the abscissae for different curves. If this fact is taken into account, some groups of films can be seen to form similar 'constellations' of points in the graphs of different wavelengths. The group of 6 points which appears at 6800 Å. between $T=0.06$ and $T=0.09$ and at 5200 Å. between $T=0.12$ and $T=0.16$ forms an example of this fact which shows that the scatter of the points in the graphs is mainly due to genuine differences of the properties of the films, and only to a small extent to errors of measurement.

Most points lie quite close to a curve showing a decrease of $R+T$ with increasing transmission, i.e. decreasing thickness of the film. The absorption $A=1-R-T$ is thus found to increase with decreasing film thickness. Except for the violet where the scatter is somewhat larger, most points lie less than $\frac{1}{2}\%$ off the curve.

The curves of Figure 2 were used in plotting R as a function of T in Figures 3 (a) and 3 (b).

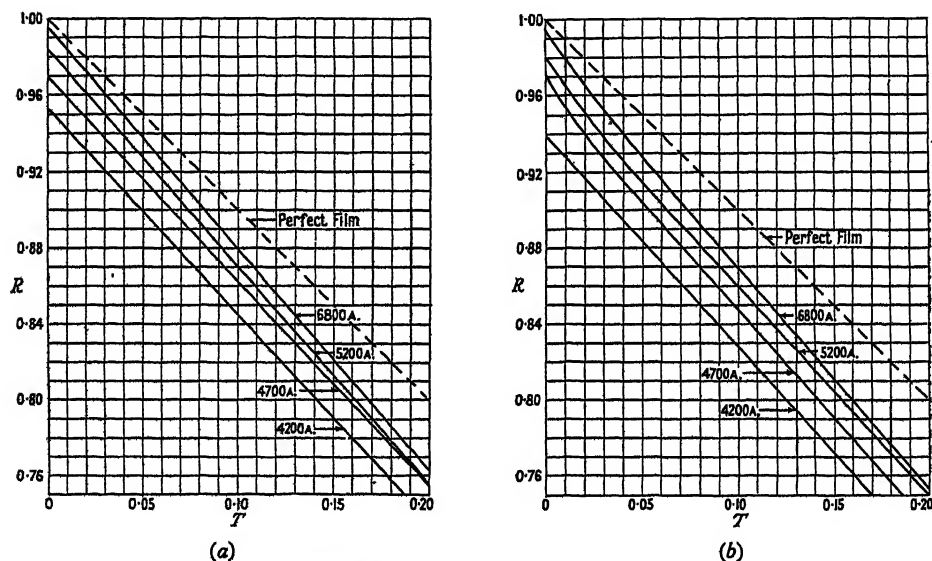


Figure 3. R as function of T .

(a) for new films, (b) for aged films (22 days).

The loss of reflectivity of the films as a result of ageing depends on the wavelength of the light and on the thickness of the film. Examination of Figure 2 and of other data shows that for dense films R decreases less rapidly than for thin films. In order to obtain a simplified picture of the process of ageing, on the basis of statistical evidence, the relative values of $R+T$ were averaged for the six films which were re-measured after 68 days, i.e. the average of $(R+T)/(R+T)_0$ for these six plates was plotted as a function of time in Figure 4, where $(R+T)_0$ is the value of $R+T$ for the new film. Since T changes little and is much smaller than R , this plot is almost identical with a plot of R/R_0 . The six films were of different thicknesses, but all of them, expressed in terms of interferometry, were fairly dense films for red, and fairly light films for violet light.

The values for red and green show that the rate of decrease of R becomes small after three weeks. The points for 4200 Å. cannot be fitted well by a simple curve;

examination of the individual results for the six plates and for others which were measured after 50 days makes it appear likely that the ageing curve for violet light is, in fact, not simple. In any case, it is certain that the reflectivity decreases with age more rapidly for shorter wavelengths.

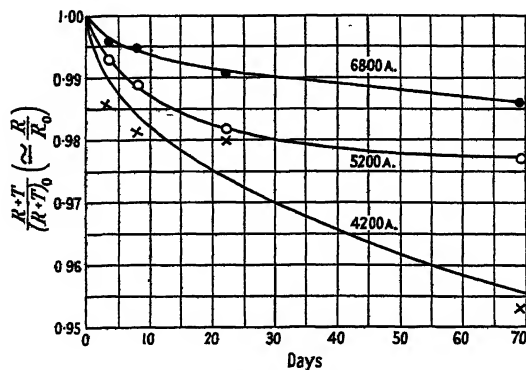


Figure 4. Relative change of R and $R+T$ with age.

A number of measurements made outside this systematic research confirms the conclusion that, in general, the rate of loss of reflectivity decreases with time; one pair of etalon plates, whose properties had not been measured accurately immediately after deposition of the films, were found to have values of $R+T$ of

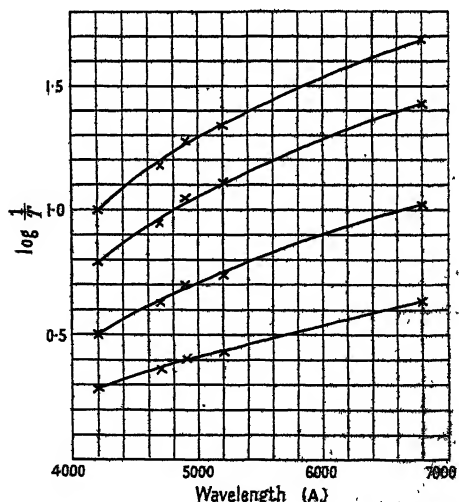


Figure 5. Optical density $D = \log 1/T$ as function of wavelength.

0.97 and 0.98 for red light after having been in use for over 6 months. In some films, the reflectivity was actually found to have increased after a few weeks, but in these cases, the initial values of $R+T$ had been abnormally low.

Figure 5 shows the optical density $D = \log 1/T$ plotted as function of the wavelength. The points forming each curve are averages of several films of very similar density.

§ 5. DISCUSSION OF THE RESULTS

Before this systematic research was undertaken, the instrument described in §2 had been used* for measuring the values of R and T for a considerable number of silver films deposited on etalon plates, most of which were made of fused silica. Most of these values agree well with the curves in Figure 2.

Since the scatter of the points in Figure 2 is much greater than the errors of the optical measurements and thus represents genuine differences in the properties of the films, it appeared desirable to establish the causes of these differences. But the large number of variables which might possibly influence the properties of the films allowed only tentative conclusions to be reached.

(i) Influence of conditions of evaporation. If vacuum conditions and speed of evaporation, within the range of variation in these experiments, had an important influence on the properties of the films, any two films deposited simultaneously at the same distance would tend to be more similar to one another than to other films of similar density. No effect of this kind was apparent. However, in comparing films of approximately equal density, but deposited at different distances from the target, some evidence was found for the smaller distance to produce slightly higher reflectivity. This effect, if real, could be interpreted as due to different speeds of evaporation.†

(ii) Influence of the type of glass. Of the 18 glass plates used, some consisted of glass appearing distinctly green when viewed end-on, while others appeared quite colourless. The results obtained with these two kinds of glass did not show any systematic differences. Silica plates, as mentioned before, also gave similar results.

(iii) Influence of surface conditions. Of six pairs of glass plates, each pair was coated together at the same distance, and the properties of the films were measured several times. The films were then removed with acid and the plates re-coated in the same way; in one case this procedure was repeated a third time. In six of these seven repeats, the plate which gave the higher value of R , compared with the 'standard' curve of Figure 2, in the first coating, also gave the higher value in the subsequent coating. The differences in question were of the order of 0.01. It thus appears that some surface conditions which are not effected by the cleaning process, are often responsible for producing a 'better' or 'worse' coating. In a few cases, imperfect cleaning was, no doubt, the cause of low values of R , and on rare occasions even of much too low values.

Previously published direct measurements of reflectivities of transparent films (Romanowa, Robzow and Pokrowsky 1934, Goos 1936, Krautkrämer 1938, Strong and Dibble 1940) show so much more scatter than the values reported here that only rough comparison is possible.

The work of Strong and Dibble (1940) covers mainly the range of low and medium reflectivity and their measurements of R are only relative and hardly accurate enough for application to interferometry. The existence of two different types of silver films below a certain thickness, as reported by these authors, may be connected with our observation that thinner films are less stable than thicker

* These measurements were carried out in conjunction with Mr. L. C. Bradley and Dr. G. W. Series, and the results were reported at the International Conference on Optical Properties of Thin Films in Marseilles in April 1949.

† Note added in proof. R. S. Sennett and G. D. Scott (*J. Opt. Soc. Amer.*, 1950, 40, 203) report a marked increase of reflectivity with increase of speed of evaporation.

ones. Our measurements do not extend to sufficiently thin films to allow direct comparison.

The measurements by Krautkrämer (1938) fall into the range of higher reflectivities, but were made with a small number of films only. Within their limits of error, stated as ± 0.01 , they are compatible with our results.

Among others, Tolansky (1946) and Dufour (1948) have given values of reflectivities, but lack of details does not allow comparison.

The reflectivity of opaque silver films, made by evaporation, was measured by Edwards and Petersen (1936). Their values for yellow and green light ($R=0.985$) are in very good agreement with our results in the limit of $T=0$ (see Figure 2).

Values of the reflectivity of silver films were derived from measurements of the width of interference fringes produced with etalons of very small spacings (Bright, Jackson and Kuhn 1949). This indirect method gave lower limits of $R+T$ which are perfectly consistent with our direct measurements. The increase of $R+T$ with increasing thickness of the films was, in these indirect measurements, disguised by the imperfections of the etalon plates, an effect which becomes more important as R increases.

For opaque films, the reflectivity can be calculated from the optical constants, but the results of such calculations have generally been found to be too low. If the values from the latest measurements of the optical constants of freshly deposited silver (Hass 1946) are compared with our limits of R for $T=0$, this discrepancy still exists. The measured values of R exceed the derived values by amounts increasing from 0.02 in the red to 0.06 in the violet.

For transparent films, the formulae of the theory are rather complicated. Calculations by Barns and Czerny (1931) show a rate of decrease of $R+T$ with increasing T which is similar to that found in our experiments, but their absolute values are consistently lower.

Though the thickness of the films was not measured, the results plotted in Figure 5 show that the intensity of the transmitted light does not vary exponentially with the thickness of the film; if the exponential law of absorption were valid, the ratio of the optical densities for any two films would be the same for different wavelengths. Figure 5 shows that these ratios are quite different for red and violet. Some indication of this fact was found by Bright, Jackson and Kuhn (1949).

One can demonstrate this effect, which is, of course, theoretically not unexpected, by a simple qualitative experiment; a pile of glass plates, each carrying a thin silver film, shows a deep violet colour in transmission, very different from the blue of a single, dense film.

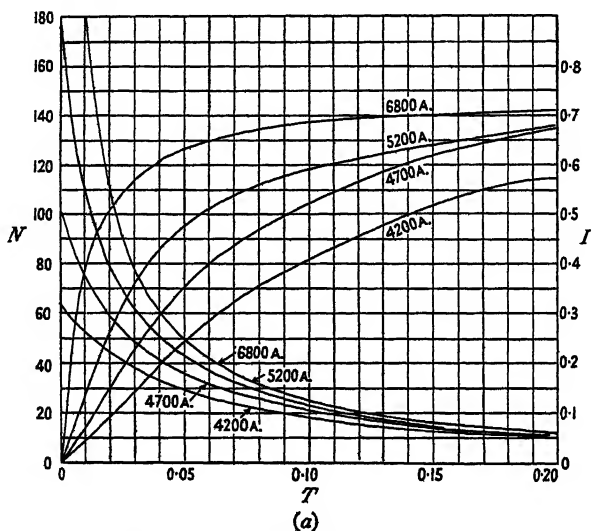
§ 6. APPLICATION TO INTERFEROMETRY

The resolving power of the Fabry-Perot etalon can be defined in various ways, and a number of authors have tabulated it as a function of R . A very convenient definition (Bright, Jackson and Kuhn 1949) considers two lines as just resolved if their distance is equal to the instrumental half-value width; the resulting intensity curve shows a depression of about 20% for lines of equal intensity, as in the Rayleigh condition for gratings and prisms. A simple expression for the resolving power can then be derived if in Airy's formula the sine of the phase angle is replaced by the angle itself, near the maxima of the intensity. This approximation is accurate enough, as long as the reflectivity is not much below 0.8. If the resolving power is expressed, in the usual way, as the product of the order of

interference n and a factor N which can be considered as the effective number of interfering beams, the latter is found to be

$$N = \pi \sqrt{R/(1-R)}.$$

With the use of the measurements described above, the value of N is plotted as a function of T , for new and for 22 days old silver films, in Figures 6 (a) and 6 (b).



5200 Å. refers to second curve, *not* first.

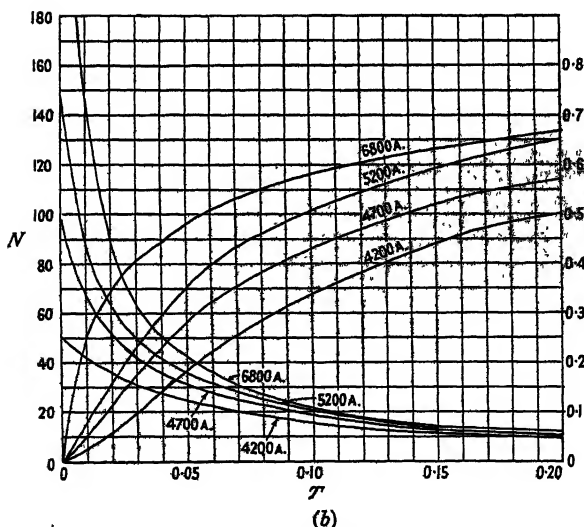


Figure 6. Effective number of beams N and intensity I of etalon fringes.
(a) for new films, (b) for aged films (22 days).

In the same Figures, the intensity of the maxima, $I = \{T/(1-R)\}^2$, is plotted. These graphs make it possible to find the value of T , for a given colour, which gives the best compromise of the requirements of resolving power and intensity. If, for example, an intensity of the fringes of 0.3 of that without any etalon is required, for a several weeks old film, Figure 6 (b) shows that for red light 1/110 of one order can be

resolved, and for green light $1/40$ of one order, provided the etalon plates are perfect. The required transmission factors are 0.015 and 0.04 respectively, for these two colours. If the value of T is to be measured with light of any other colour, the curves in Figure 5 allow the necessary conversion to be made.

It must be remembered that the formula $I = \{T/(1-R)\}^2$ gives the correct intensity only for monochromatic light. If the spectral width of the line under investigation is comparable with the instrumental width, the intensity is lower.

The surfaces of commercially obtainable etalon plates are generally not good enough to allow a smaller fraction than $1/50$ of one order to be resolved, and less than this for plates of large size. Our measurements show that, certainly for wavelengths of over 5000 Å., a considerable increase in resolving power could be achieved with ordinary silver films, if etalon plates with more perfect surfaces were available.

ACKNOWLEDGMENT

The authors wish to acknowledge the contribution due to Mr. L. C. Bradley and Dr. G. W. Series, especially in the earlier stages of this work.

REFERENCES

- BARNES, R. B., and CZERNY, M., 1931, *Phys. Rev.*, **38**, 338.
BRIGHT, R. J., JACKSON, D. A., and KUHN, H., 1949, *Proc. Phys. Soc. A*, **62**, 225.
DUFOR, CH., 1948, *Le Vide*, **3**, 480.
EDWARDS, H. W., and PETERSEN, R. P., 1936, *Phys. Rev.*, **50**, 871.
GOOS, F., 1936, *Z. Phys.*, **100**, 95.
HASS, G., 1946, *Optik*, **1**, 8.
KRAUTKRÄMER, J., 1938, *Ann. Phys., Lpz.*, (5) **32**, 537.
ROMANOWA, M., ROBZOW, A., and POKROWSKY, G., 1934, *Phys. Z. Sowjet*, **5**, 746.
STRONG, J., and DIBBLE, B., 1940, *J. Opt. Soc. Amer.*, **30**, 431.
TOLANSKY, S., 1946, *Physica*, **12**, 649.

The Absorption Spectra of Solid Lead Sulphide, Selenide and Telluride

By A. F. GIBSON

Telecommunications Research Establishment, Ministry of Supply

Communicated by R. A. Smith; MS. received 6th March 1950

ABSTRACT. Recent interest in lead sulphide, selenide and telluride as photoconductors makes information on their absorption spectra valuable. It is found that the absorption spectrum of each material is characterized by strong absorption in the visible and ultra-violet region, with a further band in the infra-red. The effects of temperature and oxidation are also studied.

In PbS the forbidden energy region between the filled and empty bands is found to be about 1.3 ev. in width, not 0.4 ev. as previously suggested. Hence photoconductivity in PbS cannot be associated with the main lattice. The corresponding energy gaps in PbSe and PbTe are found to be about 1.05 ev. and 0.9 ev. respectively.

§ 1. INTRODUCTION

THE absorption spectra of insulators and semiconductors are usually characterized by very strong absorption at short wavelengths (usually in the ultra-violet region) with a sharp edge on the long wavelength side. This absorption is generally assumed to be due to electronic transitions between the highest filled energy band and the lowest empty energy band of the solid. The absorption coefficient is usually of the order of 10^5 or 10^6 cm⁻¹.

The lattice absorption edge is often followed on the long wavelength side by an absorption band or tail of much lower absorption coefficient, of the order of 10 cm⁻¹. This absorption is generally dependent on the conditions of sample preparation and is thought to be due to impurities, cracks and other defects in the crystal lattice.

The absorption spectra of PbS, PbSe and PbTe are not exceptions to these general conditions, though the long wavelength tail absorption bands have considerably larger absorption coefficients and bandwidths than those usually encountered in sulphides.

The photoconductive properties of PbS and its associated compounds are well known (Sosnowski *et al.* 1947, Moss 1949, Chasmar 1948). The spectral response of all three materials is characterized by a sharp cut-off at long wavelengths occurring at about 3μ for PbS and about 5μ for PbSe and PbTe. In all cases the long wavelength limit moves to longer wavelengths on cooling. Early theoretical work (Sosnowski *et al.* 1947) suggested that the photoconductivity of PbS was due to absorption in the bulk lattice. This implies that the minimum energy difference between the highest filled band and the conduction band of PbS is about 0.4 ev., and less for the selenide and telluride.

The photoconductivity of PbS and its associated compounds is largely determined by suitable treatment with oxygen (Sosnowski *et al.* 1947, Schwarz 1948). Recent work in this laboratory (Gibson and Moss 1950) together with other unpublished work by Mr. A. S. Young, has shown that oxygen treatment has very marked effects on the photosensitivity spectrum. Thus in the case of PbTe photoconductivity is limited to wavelengths less than 3μ without oxygen treatment, moving to 5μ on oxidation. This and similar work suggests that photosensitivity in these materials is not primarily a property of the bulk lattice.

§ 2. EXPERIMENTAL METHODS

Lead sulphide and lead selenide layers may be prepared by two methods: evaporation *in vacuo* or chemical deposition from colloidal solution (Kicinski 1948). Layers of PbTe have not, as yet, been successfully deposited chemically and hence only evaporated PbTe layers have been examined. Layers were mounted on discs of polished quartz, artificial sapphire or KRS5 (a mixed crystal of thallium bromo-iodide) depending on the wavelength range under examination. Chemical methods of deposition gave layers of the greater reproducibility and uniformity of thickness. Within the limits of chemical analysis (1%), PbS layers deposited chemically are pure PbS.

The chief errors in absorption measurements on thin films are as follows: Firstly the layers consist of small microcrystals (of the order of 1000 Å. in size) and transmission between grains may be more important than transmission through grains. This effect is easily detected as an apparently constant absorption with wavelength and may be eliminated by careful layer preparation. In particular, chemical layers are superior in this respect to evaporated layers. A second important source of error is due to scattering of light by individual grains of the layer. This effect may be eliminated by collecting the transmitted radiation in a spherical mirror of high aperture.

The third remaining error is due to reflection. As the refractive index of PbS, PbSe and PbTe is very high (about 4.0 for PbS and over 5.0 for PbTe) the reflection coefficient is high and reflection represents a most serious source of error. When the absorption within the specimen is large (say 99% or greater) the error due to reflection can be ignored. When the percentage absorption falls to low values, however, on the long wavelength side of the lattice absorption edge, reflection introduces an apparent absorption which may be greater than the true absorption.

To eliminate reflection errors two specimen layers were prepared simultaneously under exactly similar conditions, except that one layer was arranged to be thicker than the other. The transmission through each layer was measured concurrently at each wavelength. The chief reflection, which occurs at the air-specimen interface, is the same for both layers. Errors due to reflections at the specimen-mount interface and mount-air interface are only equal in each layer if the absorption in the specimen is zero, but this is where reflection errors would otherwise be most serious. When the layers differ in transmission by a factor of five, the remaining reflection error is about 1%.

If the two layers have thicknesses d_1 and d_2 respectively, the absorption coefficient being α , then

$$J_1 = J_0(1 - R) \exp(-\alpha d_1) \quad \text{and} \quad J_2 = J_0(1 - R) \exp(-\alpha d_2)$$

where J_1 and J_2 are the intensities of the transmitted beams, J_0 the intensity of the incident beam and R the reflection coefficient of the air-specimen interface.

Hence
$$\alpha = \frac{\ln J_1/J_2}{(d_2 - d_1)}.$$

In the presentation of the results $\log J_1/J_2$ has been plotted. To obtain the absolute value of α the difference in layer thickness must be known. This is obtained by weighing, but high accuracy is not possible. Values of α are indicated alongside those of $\log J_1/J_2$.

The errors inherent in this method that are not accounted for are as follows:

(i) Interference between reflected and incident beams. This effect would lead to an apparent absorption varying cyclically with wavelength. In fact no such effect was observed. This was thought to be due to the corrugated nature of the specimen surfaces.

(ii) If both the mount and specimen absorb at any particular wavelength, then the amount absorbed by each mount is not equal. This error can be arranged to be negligible.

(iii) In spite of identical preparation conditions a pair of layers may not, in fact, be identical. Tests showed that this source of error was unimportant in chemical layers, but occasionally evaporated specimens had to be abandoned because of this feature.

Another method of eliminating reflection errors was also examined. This consisted of measuring the transmission through a 'black body' type cone of Pyrex glass coated with PbS and collecting the transmitted radiation in a spherical mirror of large aperture. The absorption curve obtained over the limited wavelength range available, (0.3μ to 3.4μ) was the same as that obtained by the difference method. The cone method was not very suitable for evaporated layers and was later abandoned.

The absorption measurements were made using a double prism monochromator with interchangeable quartz, LiF and NaCl prisms, using a thermocouple, a PbTe cell or a photomultiplier as detector. A wavelength resolution of 0.05μ or better was generally employed.

To study the absorption spectra at low temperatures both layers of each pair were totally immersed in liquid nitrogen in an unsilvered quartz dewar vessel. For some measurements layers were deposited directly on to the inner surface of the dewar vessel. The accuracy of these measurements at low temperatures was not high, say 20%, due to the constant boiling of the coolant and water condensation.

EXPERIMENTAL RESULTS

§3. THE ABSORPTION SPECTRUM OF LEAD SULPHIDE

3(i) General Results

The absorption spectrum in the range 0.2μ to 7.0μ of a typical chemically deposited layer of PbS is given in Figure 1. If care is taken to use a standard deposition method the layers are quite reproducible. Deviations from the standard method cause small changes in the absorption spectrum, particularly around 0.9μ . Layers may be prepared with a very marked dip at this wavelength which separates the lattice absorption and the tail into two distinct bands.

Evaporated layers of PbS give very similar absorption spectra to those of chemical layers. The chief differences occur (a) in the long wavelength tail band which may differ in magnitude by as much as a factor of 3 and (b) in the wavelength region around 0.6μ . The original PbS available for the preparation of evaporated layers is, however, of doubtful purity. Thus two samples of 'pure PbS' available commercially have been found to contain 45% and 55% of PbSO_4 respectively. The chemically deposited layers probably represent the purest specimens available.

The chief characteristics of the absorption spectrum of PbS are easily distinguished, the lattice edge being very marked. The absorption coefficient in this region is about $2 \times 10^5 \text{ cm}^{-1}$, which is a reasonable value for any semiconductor (Mott and Gurney 1940). The most surprising feature of the absorption spectrum is the marked tail band which extends beyond 6μ . The absorption coefficient in this band is about 10^4 cm^{-1} and is therefore much larger than that found in semiconductors. The significance of this result will be discussed later.

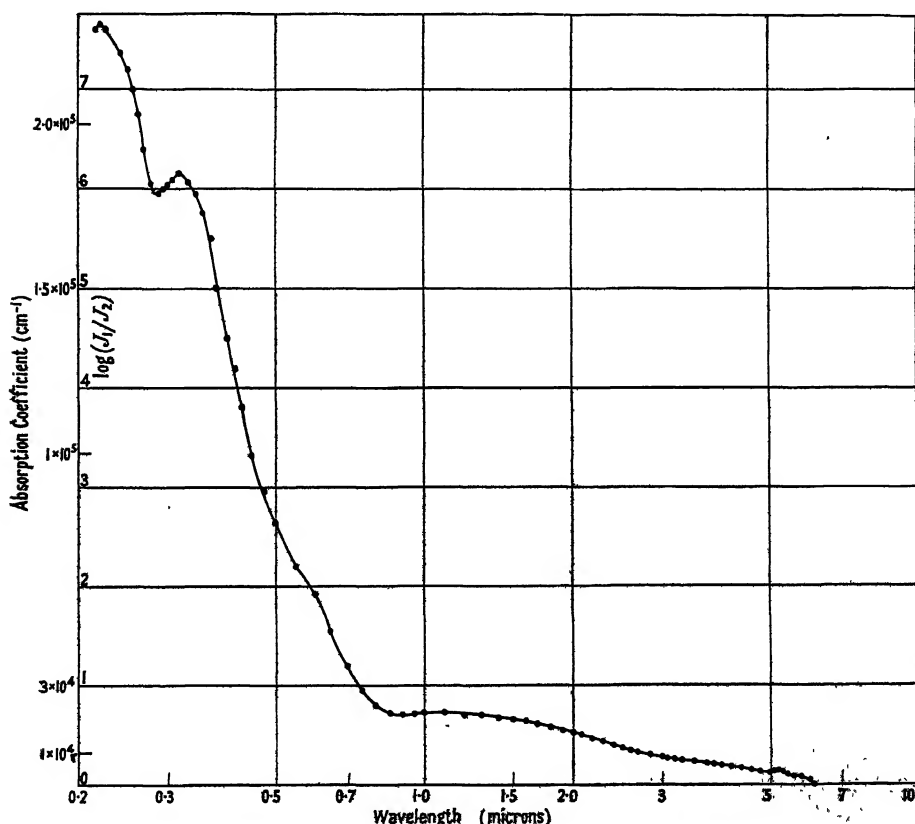


Figure 1. The absorption spectrum of chemically deposited lead sulphide.

3 (ii) The Effect of Oxygen Impurity

The importance of treatment with oxygen to obtain photosensitive layers of PbS is well known.

Oxygen may be introduced either (a) during the preparation of the original material, or (b) during evaporation, or (c) after the layer has been deposited. Generally speaking only processes (b) and (c) have marked effects on the photosensitivity.

Chemical layers of PbS are normally baked in air for about 30 minutes at about 150°C . in order to produce photosensitivity. This corresponds to oxidation process (c) above. Initial measurements indicated that such treatment made no significant change in the absorption spectrum. Bakes up to 16 hours

duration were tried. This result was not altogether unexpected as von Hippel *et al.* (1946) have reported the same result for thallium sulphide, another material normally sensitized by baking in air or oxygen.

A significant effect due to oxygen was observed when care was taken to reduce all extraneous effects to a minimum. In particular the layers were not removed from the monochromator mounting for baking. Evaporated layers were prepared *in vacuo* and their transmission measured at each wavelength before air at atmospheric pressure was admitted by breaking a small glass seal.

The change in absorption observed was very small (about 3% or less) but almost doubled on baking in air. The magnitude of the change was wavelength dependent, being most marked at the wavelengths 1.0μ and 2.3μ in the tail band. In the tail absorption band the absorption decreased on admission of oxygen.

In view of the small change in absorption to be observed the errors in this measurement are large, but the reality of the effect is beyond doubt. The most surprising feature is that the absorption coefficient in the tail absorption band decreases on oxidation, when an increase would be expected if new impurity bands are being introduced. It may be significant that previous workers (Sosnowski *et al.* 1947) have noticed peaks in the photoconductivity spectrum of PbS at about 1μ and 2μ which they ascribe to oxygen treatment. These wavelengths agree fairly well with the figures given above.

It has been stated by Wilman (1948) and others that as much as 10% of PbO.PbSO_4 may be formed on the oxidation of PbS films. It is almost certain that the main absorption band of this material lies in the far ultra-violet outside the range of measurement.

3 (iii) *The Effect of Temperature*

The absorption spectrum of PbS at room temperature and at 77°K . is shown in Figure 2, the abscissa being the reciprocal of the wavelength. The most obvious effect of cooling is to shift the lattice absorption edge to longer wavelengths. A number of determinations of the magnitude of the shift have been made, the mean value indicating a shift of about $6 \times 10^{-4}\text{ eV}/^\circ\text{K}$. This compares with the measured shift in the spectral limit for photoconductivity of $4.8 \times 10^{-4}\text{ eV}/^\circ\text{K}$, and is in the same direction (Moss 1949).

A lattice edge shift to longer wavelengths on cooling is not usual in semiconductors. There are, in general, two factors which determine the shift of the lattice edge with temperature. As the lattice contracts on cooling ions move closer together, interaction increases and allowed energy bands broaden. In addition, however, electron interaction with the thermal vibrations of the lattice decreases on cooling and allowed energy bands narrow (Radkowsky 1948). It is not possible without detailed calculation to determine which effect will be the more important in PbS. The energy change due to contraction is very approximately proportional to energy bandwidth, whereas the shift due to lattice vibration effects is independent of bandwidth. As the former effect appears to be the more important in PbS it is concluded that the allowed energy bands in PbS are very wide. In general this would be expected in a sulphide of high dielectric constant.

The long wavelength tail absorption band of PbS also undergoes a change on cooling. The band appears to narrow appreciably, at least on the long wavelength side, and the peak of the band moves to longer wavelengths. The

total area under the tail band is not appreciably altered. It should be pointed out that measurements at 77° K. could not be made at wavelengths greater than 4.2 μ because of absorption in the quartz dewar vessel. In addition measurements around 2.7 μ were impossible due to the condensation of water on the cooled surfaces of the dewar vessel. Thus the accuracy in the 3 μ and 4 μ region is not very high.

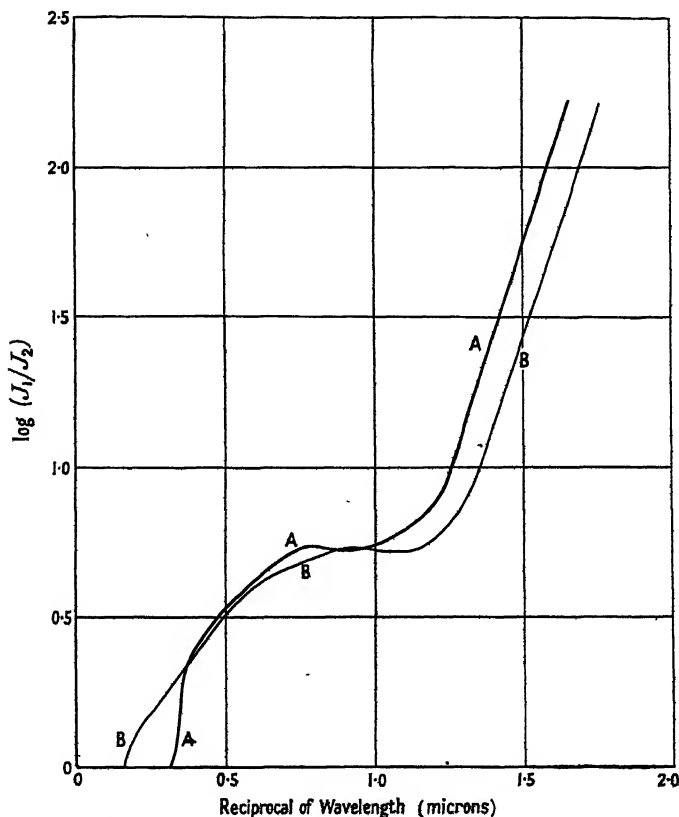


Figure 2. The absorption spectrum of lead sulphide.
A at 77° K., B at 290° K.

3 (iv) Other Measurements on PbS

(a) In view of the lack of marked effects due to the exposure of evaporated PbS films to the air, layers were prepared *in vacuo* on KRS5 discs and subsequently opened to the air. As KRS5 is transparent to very long wavelengths (to about 40 μ) it was possible to examine the absorption of PbS out to 16 μ (the limit set by the monochromator and detector). No major absorption bands in the region 7 μ to 16 μ were observed. The measurements set an upper limit to the 'Reststrahlfrequenz' in PbS.

(b) Attempts have been made to measure the absorption spectra of single crystals of natural galena. These experiments were not very successful as it was impossible to obtain crystal specimens sufficiently thin and yet free from holes. The measurements obtained indicated a similar structure to that shown in Figure 1, the lattice edge being easily observable.

Due to the above difficulties it was not possible to observe changes in absorption with crystal direction by transmission measurements. As an alternative method crystals were split and polished and mounted on a table rotatable through 270° . The reflectivity of the specimens was then measured with plane-polarized infra-red radiation at various crystal angles. No significant changes in reflectivity with crystal direction were observed.

§4. THE ABSORPTION SPECTRUM OF LEAD SELENIDE

The absorption spectrum of a chemically deposited layer of this material is given in Figure 3 where the absorption coefficient is plotted against the

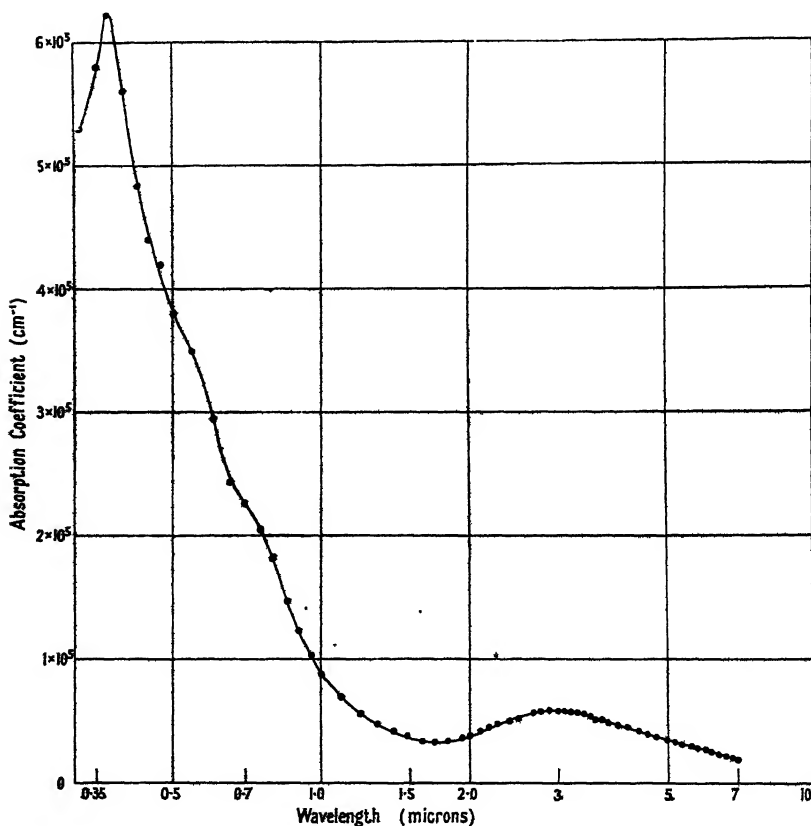


Figure 3. The absorption spectrum of chemically deposited lead selenide.

logarithm of the wavelength. The spectrum is similar to that of PbS except that it is shifted bodily to longer wavelengths. The shift is about 0.4 ev. in the region of lattice absorption.

The lattice edge of PbSe, like that of PbS, moves to longer wavelengths on cooling. The shift is about the same as, or rather less than, that of PbS. The expansion coefficients of PbS and PbSe are almost identical but the allowed energy bands are likely to be somewhat narrower in PbSe. According to Moss (1949) the shift in spectral limit of photoconductivity of PbSe is exactly the same as that of PbS.

§ 5. THE ABSORPTION SPECTRUM OF LEAD TELLURIDE

This material cannot be obtained by chemical deposition and hence only evaporated films may be studied. PbTe can, however, be prepared by heating the elements in sealed and evacuated quartz tubes and hence a product of high purity (particularly as regards oxygen) obtained. This represents an important advantage over PbS and PbSe.

The absorption spectrum of a layer of PbTe prepared by the evaporation *in vacuo* of a powder sample manufactured *in vacuo* is given in Figure 4. The general features are similar to those of the sulphide and selenide.

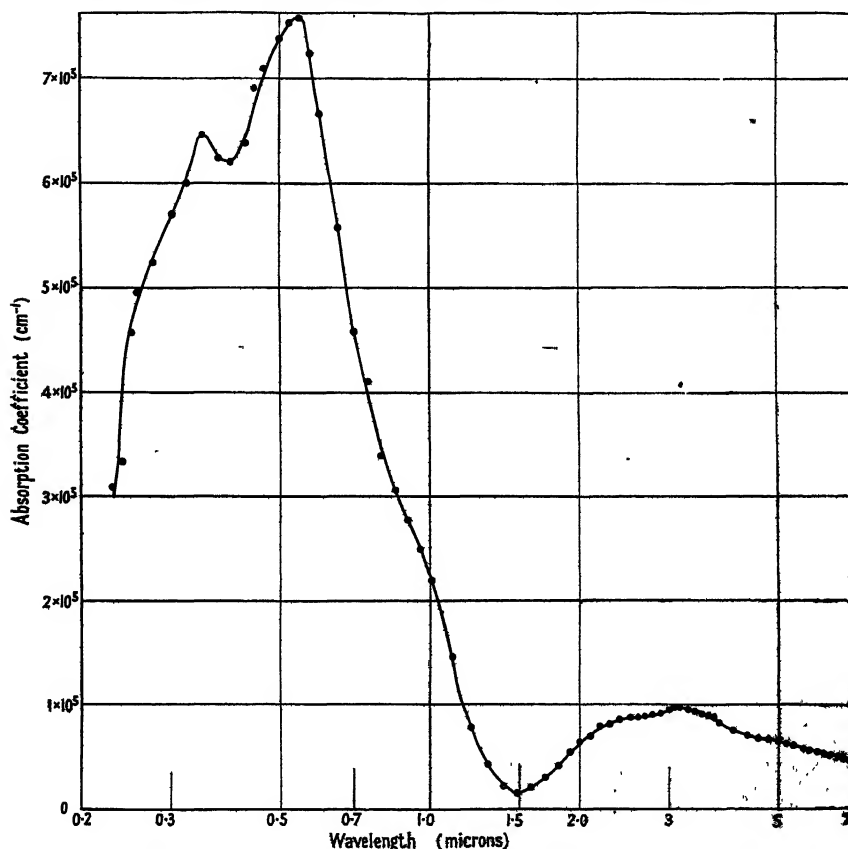


Figure 4. The absorption spectrum of lead telluride (material prepared in high vacuum).

The effect of oxygen on the absorption spectrum of this material can be studied both in the bulk material and in films. As for PbS, the absorption in the long wavelength tail band decreases slightly when air is admitted to vacuum-prepared films but the effect is smaller than that in the sulphide and too small for adequate measurement.

To study the effect of oxygen on the bulk material, specimens of PbTe have been prepared by heating the elements in low pressures of air and in open crucibles. The effect of such treatment is illustrated in Figures 5 and 6. The absorption at short wavelengths is not appreciably affected (the main peak at $0.55\ \mu$ remains entirely unchanged in position or magnitude) but marked changes occur at longer wavelengths. A new absorption peak appears at about $0.9\ \mu$ and the

tail absorption decreases in magnitude on its short wavelength side. The latter effect results in an apparent shift to longer wavelengths of the tail band maximum and the minimum between the main bands.

This comparatively violent oxidation (the material examined in Figure 6 was heated to over $1,000^{\circ}\text{C}$. in air) is clearly similar, except in magnitude, to oxidation of sulphide and telluride films by exposure to the air. Similar though less marked effects can be obtained by evaporation of unoxidized PbTe samples in low pressures (of the order of 10^{-2} mm.Hg) of air. It seems likely that all

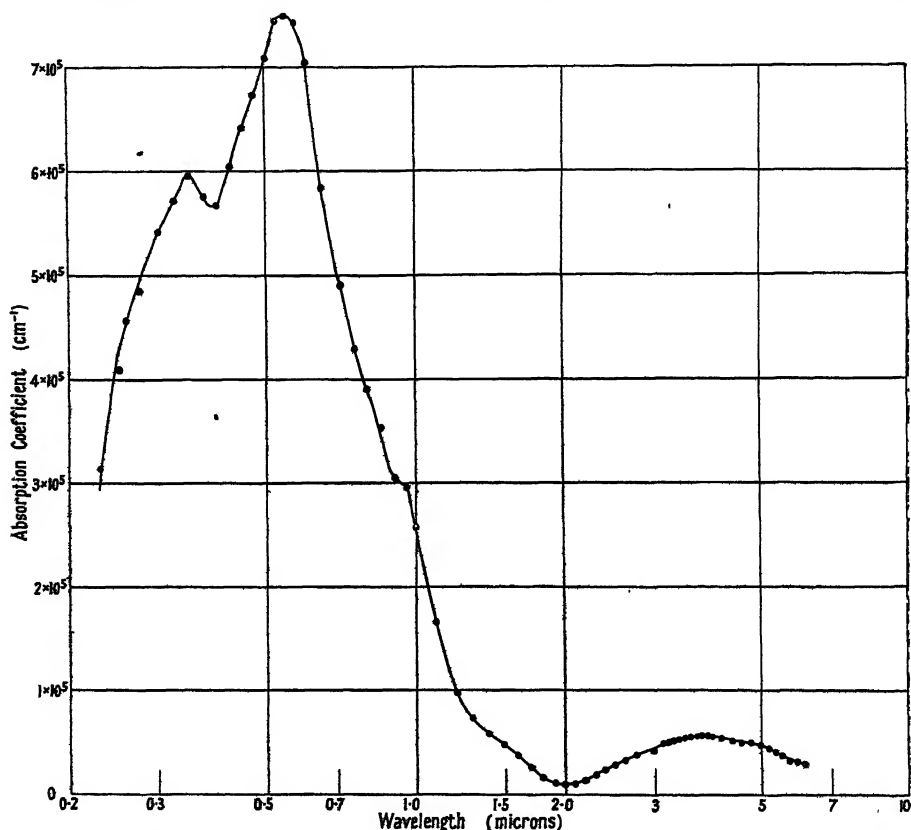


Figure 5. The absorption spectrum of lead telluride (material prepared in low pressure of air).

the more marked differences in absorption spectra between PbS, PbSe or PbTe samples of different origin are due to oxidation effects.

Very intense oxidation of PbTe results finally in complete decomposition and the formation of a grey-white product. By analogy with PbS, the material is thought to be $\text{PbO} \cdot \text{PbTeO}_4$. The only marked absorption of this material commences about 2800 \AA . and extends below the limit of measurement. It seems likely that the absorption bands of $\text{PbO} \cdot \text{PbSO}_4$ would be at even shorter wavelengths.

The shift in the lattice absorption edge in PbTe on cooling is even less than that of the selenide and its magnitude proportionately uncertain. A reasonable value would appear to be a $1.5 \times 10^{-4} \text{ eV}/^{\circ}\text{K}$. According to Moss the shift in photoconductive limit is the same as for the sulphide and selenide but recent work in this laboratory by Mr. A. S. Young has shown that any shift between 2 and $7 \times 10^{-4} \text{ eV}/^{\circ}\text{K}$. may be obtained in PbTe, depending on sample preparation.

§ 6. THE ABSORPTION SPECTRA OF PbS, PbSe AND PbTe: GENERAL DISCUSSION

6 (i) Lattice Absorption

The lattice absorption bands of each material almost certainly correspond to electronic transitions between the highest filled energy band and the lowest empty band, that is, the transfer of an electron from S, Se or Te ions to Pb ions. The long wavelength part of this band may correspond to exciton transitions (Mott and Gurney 1940), but there is no evidence in favour of this view.

Extrapolation indicates that the lattice absorption coefficient of PbS will reach zero at about 0.9μ , corresponding to a minimum energy of about 1.3 or 1.4 eV. This value is considerably greater than the value of 0.4 eV. suggested by Sosnowski *et al.* (1947) and suggests that these authors were mistaken in

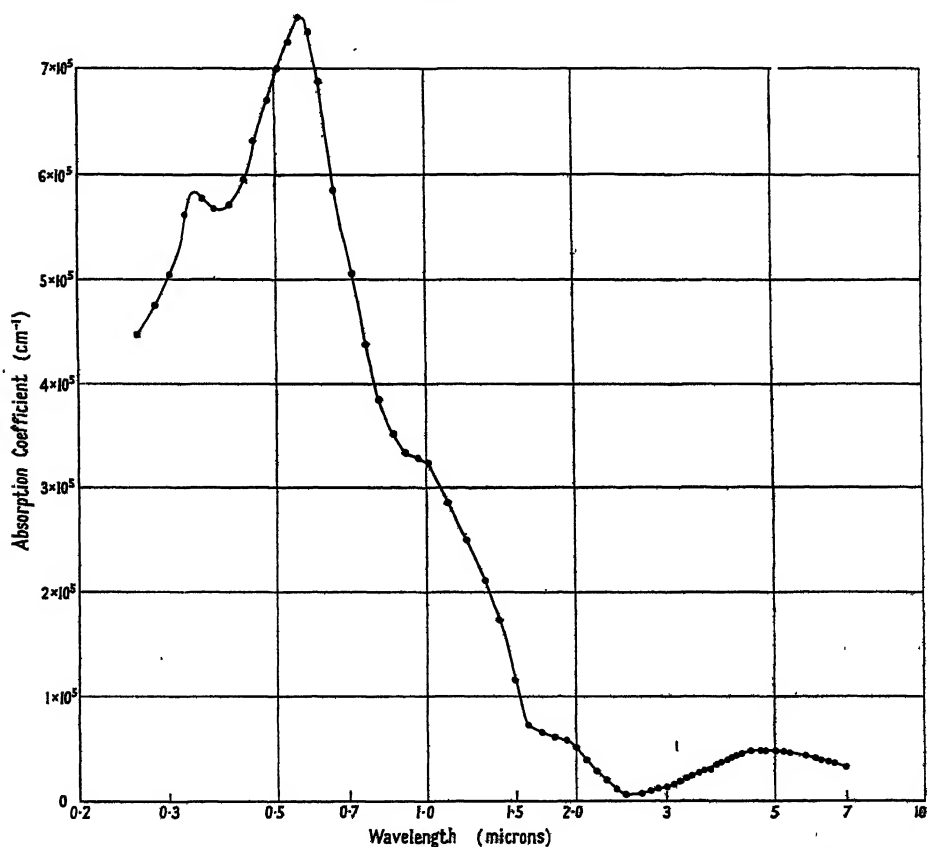


Figure 6. The absorption spectrum of lead telluride (material prepared in air).

associating photoconductivity with lattice absorption. It would appear that the photoconductive spectrum is in fact determined by impurities or other effects.

It is of interest to note that the sensitization spectrum of optically sensitized PbS layers (Gibson 1949) is almost exactly coincident with the main lattice absorption. Previously sensitization was thought to take place at specks of free sulphur but the similarity between the sensitization spectrum and the photoconductivity of free sulphur is not very close. It is now clear that a more likely

explanation of optical sensitization involves absorption in the main lattice and the subsequent filling of electron traps lying just below the conduction band of the solid.

The threshold energy for lattice absorption falls progressively from the sulphide to the telluride, as might be expected. The magnitude of the absorption coefficient increases through the series. The latter effect must be associated with decreasing energy bandwidths as the spectra move to longer wavelengths. Due to this progression almost all the main absorption band of PbTe lies within the limits of measurement and an estimate of the bandwidth may be made. The half width of the absorption band is about 4 e.v. which must be approximately equal to the sum of the widths of the two allowed energy bands involved.

6(ii) *The Long Wavelength Tail Absorption Band*

By analogy with other photoconductors and semiconductors this band may be ascribed to impurities in interstitial or substitutional lattice positions. There are, however, a number of reasons for rejecting this explanation, as follows.

(a) The absorption coefficient is too large, being about 10^4 cm^{-1} in PbS and larger in PbTe. Absorption bands of the magnitude and width observed would correspond to an impurity concentration of the order of 10^{20} per cm^3 . PbS, PbSe and PbTe are highly polarizable materials and therefore it may be argued that they are capable of taking up very large amounts of impurity without serious lattice distortion. Notwithstanding this, it seems unlikely that a chemically deposited film of PbS, which is never heated, should contain about 100 times as many impurities as are usually found in other sulphides.

(b) If photoconductivity is associated with impurities and the excitation of electrons from impurity levels, the absorption bands observed cannot correspond to such transitions, as the absorption spectra and photoconductivity spectra are not the same. At room temperature absorption in PbS extends to 6 or 7 μ , photoconductivity to 3 μ .

An alternative explanation is to ascribe the observed absorption band to electrons in the conduction band of the solid, the transition involved being to the next highest empty band, which probably overlaps the conduction band. The measurements of Eisenmann (1940) and others suggest that the concentration, of electrons in the conduction band of PbS is of the order of 10^{19} or 10^{20} at room temperature. This number appears to be largely invariant with temperature (Dunaer 1947). Thus, PbS, PbSe and PbTe are semi-metallic (Chasmar 1948). These electrons would give rise to an absorption band at long wavelengths of about the magnitude observed. The area under the band would be largely unaltered on cooling, though the energy distribution of the electrons may alter. No photoconductivity would arise from this type of transition and hence the lack of photoconductivity at 6 μ is accounted for. The effect of oxygen in diminishing the absorption band may be interpreted as the capture of electrons by oxygen atoms on the surface of the crystallites.

If this explanation is correct, photoconductivity cannot arise from an increase in the number of electrons in the conduction band as the efficiency would be too low. Schwarz (1949) has suggested that photoconductivity is a surface phenomena arising from adsorbed films of oxygen. A theory of this type has been considered for some time in this laboratory. Each crystallite of PbS (n-type) is assumed to have oxygen atoms adsorbed on its surface which capture electrons from the

interior of the crystallite and produce a negative surface charge. Illumination may eject electrons from the surface oxygen states, the surface charge would be reduced and conduction facilitated. There is increasing experimental evidence in favour of this view but conclusive evidence is still lacking.

ACKNOWLEDGMENTS

I am indebted to my colleagues at the Telecommunications Research Establishment for much help, advice and criticism, and particularly to those to whom reference has been made in the text. I am indebted to the Chief Scientist, Ministry of Supply and the Controller, H.M. Stationery Office, for permission to publish this work.

REFERENCES

- CHASMAR, R. P., 1948, *Nature, Lond.*, **161**, 281.
DUNAER, J. A., 1947, *C. R. Acad. Sci., URSS.*, **55**, 21.
EISENMANN, L., 1940, *Ann. Phys., Lpz.*, **38**, 121.
GIBSON, A. F., 1949, *Nature, Lond.*, **163**, 321.
GIBSON, A. F., and MOSS, T. S., 1950, *Proc. Phys. Soc. A*, **63**, 176.
VON HIPPEL, A., *et al.*, 1946, *J. Chem. Phys.*, **14**, 355, 370.
KICINSKI, F., 1948, *Chem. and Ind.*, **17**, 54.
MOSS, T. S., 1949, *Proc. Phys. Soc. B*, **62**, 741.
MOTT, N. F., and GURNEY, R. W., 1940, *Electronic Processes in Ionic Crystals* (Oxford: Clarendon Press).
RADKOWSKY, A., 1948, *Phys. Rev.*, **73**, 749.
SCHWARZ, E., 1948, *Nature, Lond.*, **162**, 614; 1949, *Proc. Phys. Soc. A*, **62**, 530.
SOSNOWSKI, A., *et al.*, 1947, *Nature, Lond.*, **159**, 818.
WILMAN, H., 1948, *Proc. Phys. Soc.*, **60**, 117.

Note on the Electron Velocity Distribution in Low Voltage Arcs

By M. S. B. CHAGHTAI

Queen's University, Belfast

Communicated by K. G. Emeléus; MS. received 20th March 1950

ABSTRACT. An analysis has been made of a large number of hot cathode low voltage arc-type discharges in mercury vapour and hydrogen at low pressure by the cold probe methods of Langmuir and Mott-Smith. It has been verified that the space potential in the plasma near the filament is positive relative to anode potential, and it has been shown that consistent data for the space potential and concentrations of the slow (ultimate) electrons are obtainable from analyses of probe characteristics in both retarding and accelerating fields for electrons. The way in which the fast (primary) electrons develop a random motion has been traced by progressively confining them within the main discharge, firstly by using a cylindrical gauze anode with open ends, secondly by using a solid anode with open ends and lastly by using a solid anode with end caps.

§ 1. INTRODUCTION

IN an hot cathode arc discharge at low pressure electrons from the cathode are accelerated through a double space-charge sheath and produce a highly conducting plasma further out. The collision processes occurring in the plasma have been investigated in detail by Langmuir and Jones (1928), who have studied how the primary electrons have their directions altered and give up energy in various types of collisions with the gas molecules. They were able to distinguish three main groups of electrons in the plasma; the primary electrons from the filament, the secondary electrons ejected from atoms and produced by loss of energy of the primary electrons, and the slow 'ultimate' electrons which are mainly responsible for the high conductivity. The tube used had a straight filament cathode mounted axially in a closed cylindrical metal box, with the end caps insulated from the curved wall. By varying the gas pressure, Langmuir and Jones made an investigation* of the transition from the velocity distribution principally studied, which can be approximately described by the three groups, to the velocity distribution which is encountered under many other conditions (Langmuir and Mott-Smith 1924), in which a single group is present, often with a Maxwellian distribution of velocities. The present author has found that some simple modifications of the electrode system used by Langmuir and Jones permit the demonstration of the transition in another way. The modifications consist in the use of different types of anode, which progressively confine the faster electrons more within the general body of the arc plasma. This shows rather directly how the overall velocity distribution tends to become Maxwellian, although both the atomic process involved and the current voltage characteristics of the collectors are too complex to permit full theoretical analysis.

* See particularly Figure 13 of their paper.

§ 2. APPARATUS AND EXPERIMENTAL METHODS

Most of the experiments were performed in mercury vapour saturated at a temperature of approximately 20°C . A few experiments performed with hydrogen at the same pressure gave similar results. No allowance has been made for thermal effusion. The tubes had an axial tungsten filament cathode 0.2 mm. in diameter, used as a bright emitter. The length of the filament between its supporting leads varied from 6.0 to 6.5 cm. as replacements of it were made. The fall of potential along the filament was approximately 8 volts. The anode of the first tube (Figure 1) was made of nickel gauze (with 25 wires/cm., each wire 0.15 mm. diameter) of length 4.2 cm. and diameter 2.3 cm. In the second tube the nickel gauze was replaced by nickel sheet of the same dimensions, and in the third tube the nickel sheet was furnished with two flat end caps through which the filament and probe leads passed. Two cylindrical molybdenum probes were used. The first F_1 was situated at about the middle of the filament, and the second F_2 at one end in such a position that its exposed length always lay within the perpendicular from the corresponding end of the anode on the filament. Both probes were 0.1 mm. in diameter, F_1 being 0.8 cm. long and F_2 0.4 cm. long. They were

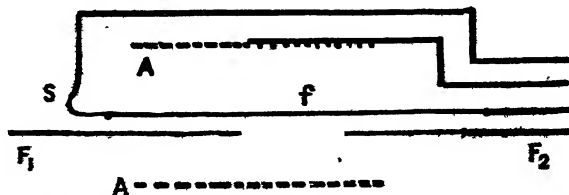


Figure 1. Diagram of electrodes and leads for tube 1. Gauze anode, AA; filament, f, held taut by sprung metal strip, S, at left-hand end; probes, F_1 and F_2 , with glass shields omitted. The four leads on the right passed to a common pinch. The lead to F_1 passed through a separate seal at the left.

shielded by fine glass tubes except at the exposed ends. The probes were parallel to the filament, and about 3 mm. from it. They were put close to the filament in order to minimize complications from plasma oscillations, which do not usually become pronounced until farther out (Neill 1949). By reversing the direction of the current through the filament, information could be obtained about the discharge at four different regions along its length from the negative end. The electrodes were mounted axially in a glass tube 4.5 cm. in diameter and 15 cm. long, with an appendix containing a drop of pure mercury. The tube was kept continuously evacuated by a diffusion pump.

Most measurements were made with the tubes passing temperature-limited currents from the whole of the filament. A few were made with the current not quite saturated. No significant differences were found between the two sets of results. With tubes 1 and 2 the plasma could be seen to bulge beyond the open ends of the anode, in some instances filling the whole of the tube. With tube 3 there was occasionally a small penetration of plasma past a filament lead. The characteristic curves were obtained using standard circuits as described, for example, by Greeves and Johnston (1936), and runs in which variation of probe potential caused a detectable change in anode current were rejected. The general shape of the characteristic curves was that described by Langmuir and Jones. For sufficiently negative voltages of the probes a small positive ion current was received which did not vary much with voltage. When the probe voltage was

raised to that of the neighbouring part of the filament a rapid change in current occurred, due to the reception of primary electrons. With increasing positive voltages more primary electrons were received. On approaching the anode potential a further rapid increase of electron current to the probe occurred, due to the reception of the ultimate electrons. In no case was it possible to distinguish clearly between primary and secondary groups.

§3. ANALYSIS OF RESULTS

As the probes were situated close to the filament it was essential to know that they were actually in the plasma. This was considered likely, from the appearance of the discharge and from the fact that the cathode sheath was probably much nearer the cathode (Druyvesteyn and Penning 1940). It was confirmed by the analysis of the electron currents received by the probes when their potential was near that of the anode. Since the ultimate electrons were present in much greater numbers than the primary electrons, the latter could be allowed to pass and the current of the probe characteristic by approximate graphical extrapolation. The currents of ultimate electrons were analysed in two ways, first, by plotting the currents on a semi-logarithmic scale against the probe voltage (i.e. $(\log i, V)$ graph), and second, by plotting their squares against the probe voltage, on a linear scale (i.e. (i^2, V) graph). Both methods permit the determination of the space potential and the electron concentration (Langmuir and Mott-Smith 1924). In the second method it is necessary to apply two corrections: first, for the position of the space potential, using a value for the mean electron energy obtained by the first method, and second, for the distortion of the linear $(\log i, V)$ graph through incomplete orbital motion near the space potential. Reasonably consistent results were obtained by the two methods, as is shown by some typical data given in the Table. In all cases the space was positive with respect to the anode.

Plasma Data

Tube and probe	Pressure (10^{-3} mm.)	Arc current (ma.)	Space potential (volts)		Ultimate electron concentration ($10^9/\text{cm}^3$)	
			from $(\log i, V)$	from (i^2, V)	from $(\log i, V)$	from (i^2, V)
1, F_1	1.35	160	7.5	7.4	25.6	20.9
1, F_2	1.35	176	7.0	7.1	0.3	0.2
2, F_1	1.24	220	5.0	5.0	21.2	18.8
2, F_2	1.24	200	2.2	2.0	1.1	0.8
3, F_1	1.04	347	4.5	4.5	18.6	16.8
3, F_2	1.04	340	1.7	1.7	0.4	0.4

The arc voltage drop was 25, and the filament current 4.85 amp. in each case.

The distribution analysis outlined above suggested that if the primary electrons were progressively confined within the inter-electrode space by partial reflection from the surface of suitably designed anodes, the overall distribution would tend to become more nearly Maxwellian. Tubes 1, 2 and 3 (see §2) were designed and used one after the other with this end in view. In tube 1 a number of the faster electrons were being lost to the main body of the plasma because they passed out through the wires of the gauze. In tube 2 there was no penetration of the anode by fast electrons, although not all were reflected from the metal (Langmuir and Jones 1928). In tube 3, loss of electrons through the ends of the anode was prevented in addition. Results of the type anticipated were obtained.

Consideration was given to the possibility that the space potential was near the cathode potential, corresponding to a point on the $(\log i, V)$ graph at which the rate of increase of primary electron current with increase of probe voltage begins to diminish rapidly. This alternative interpretation of the $(\log i, V)$ graph has been rejected from the close analogy between the discharges studied in this investigation, and those studied by Langmuir and Jones (1928) and Compton and Eckart (1925) in which the space potential was near that of the anode.

An attempt has been made to determine the velocity distribution amongst the primary electrons from analysis of the $(\log i, V)$ curves with the probe potential not far from that of the cathode. Qualitatively the distribution, particularly for tube 1, clearly consists of a large component of drift motion with a superimposed

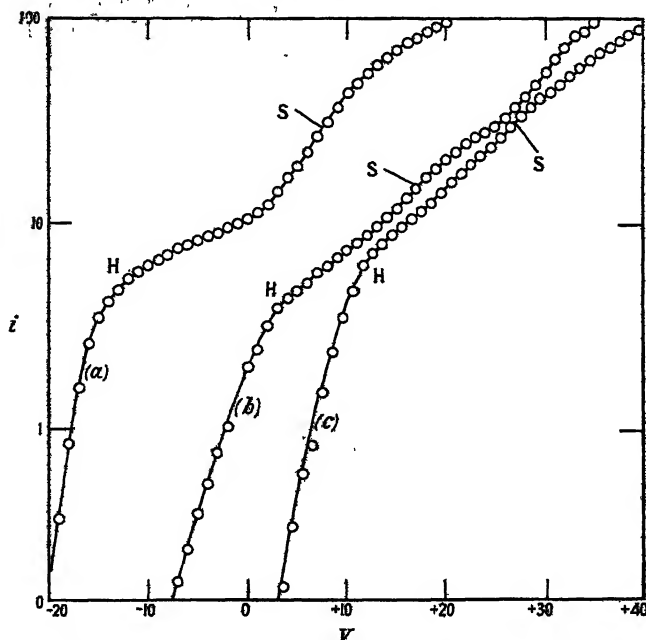


Figure 2. Semi-logarithmic plots of electron current i on arbitrary scale against probe voltage V relative to anode. Curve (b) is displaced 12.5 volts to the right and curve (c) 22.5 volts to the right. The space potentials are at S.

smaller random motion. Methods for analysing such curves have been developed by Langmuir and Mott-Smith (1926). Three complete semi-logarithmic plots of electron current against probe voltage are shown in Figure 2, for tube 1 (curve *a*), tube 2 (curve *b*) and tube 3 (curve *c*). The discharge conditions were: for curve *a*, pressure 1.35×10^{-3} mm., arc current 160 ma.; for curve *b*, pressure 1.24×10^{-3} mm., arc current 220 ma.; for curve *c*, pressure 1.04×10^{-3} mm., arc current 347 ma. In each case the arc voltage drop was 25.0, and the filament current 4.85 amp. Further data are included in the Table. The curves show a transition from the form in which the primary electrons are well separated (*a*) towards a form in which the semi-logarithmic plot is more nearly straight over the greater part of its length negative to the space potential. This type of transition has been found consistently in some 100 characteristic curves for F_1 ; it occurs in all cases by simultaneous flattening of the left-hand hump H, and extension towards more negative voltages

of the approximately straight part corresponding to the ultimate electrons near the space potential. Since the existence of a straight-line curve for $(\log i, V)$ shows that the electrons concerned have a Maxwellian distribution of velocities, we appear to have here an example of the degeneration of the separate groups towards a single group having an overall Maxwellian distribution and equivalent temperature. This is what would be expected through the progressive enclosure of the discharge, with diminishing leakage from the cylinder of primary electrons, and, possibly to a less extent, of ultimate electrons. The characteristic curves have proved to be too complex to permit quantitative analysis, but it does appear that not only is there a modification of the primary group, but in addition, an extension towards higher energies of the ultimate group, and that, could the discharge conditions be varied widely enough to follow these processes still further, a single group with a definite temperature (much above the gas temperature because of the presence of electric fields) would be obtained.

Confirmation of the essential correctness of this interpretation of the data for F_1 is afforded by the results obtained with F_2 . When this was at the negative end of the filament, the characteristic curves for tubes 1 and 2 were similar to one another, whilst with tube 3 only a slight straightening occurred. The similarity of the characteristic curves for tubes 1 and 2 is due to the magnetic field of the filament current. This bends the electron beam emerging from the filament sheath into the cylinder, and therefore robs the region surrounding this end of a considerable number of primary electrons. This bending occurs with both tubes, and although the laws of electron reflection are unknown in detail and probably vary considerably, the increased reflection from the solid anode in tube 2 can scarcely result in the redirection of many electrons towards the negative end. Hence the overall velocity distribution there remains practically unchanged. In tube 3, again, little alteration of the characteristic curves would be expected because the axial component of the velocity of the primary electrons is small.

When F_2 was at the positive end of the filament the probe characteristics were found to change in type much as they did with F_1 , for all three tubes. This again is to be expected, because at this end the magnetic field tends to make the primary electrons pass out of the cylinder.

ACKNOWLEDGMENTS

The experiments described are part of an investigation of low voltage arcs made by the author whilst a research student of the Government of Pakistan at Queen's University, Belfast. The author wishes to express his thanks to the Pakistan Government, and to Professor K. G. Emeléus, in whose laboratory the work was done.

REFERENCES

- COMPTON, K. T., and ECKART, C., 1925, *Phys. Rev.*, **25**, 139.
 DRUYVESTEYN, M. J., and PENNING, F. M., 1940, *Rev. Mod. Phys.*, **12**, 87.
 GREEVES, F. D., and JOHNSTON, J. E. McF., 1936, *Phil. Mag.*, **21**, 659.
 LANGMUIR, I., and JONES, H. A., 1928, *Phys. Rev.*, **31**, 357.
 LANGMUIR, I., and MOTT-SMITH, H. M., 1924, *Gen. Elect. Rev.*, **27**, 449 etc.; 1926, *Phys. Rev.*, **28**, 727.
 NEILL, T. R., 1949, *Nature, Lond.*, **163**, 59.

The Determination of the Viscosity of Molten Metals

By M. R. HOPKINS AND T. C. TOYE

The British Iron and Steel Research Association, Swansea

MS. received 16th February 1950

ABSTRACT. A method is described for the measurement of the viscosity of molten metals at temperatures up to about 900° C. The metal is contained in a cylindrical crucible which is allowed to oscillate about its axis under the control of the torsion in a suspension wire. The theory of the method is considered.

The viscosity of zinc has been determined in the temperature range 430° C. to 480° C. and the results are given.

§1. INTRODUCTION

THERE is not a great deal of information available on the coefficients of viscosity of molten metals. Still less is known of the viscosities of alloys and the effect, upon the viscosity of one metal, of the addition of small quantities of another.

Most of the well-known methods for the determination of the viscosity of substances which are liquid at room temperature are unsuitable for measurements on liquid metals. The use of the oscillating disc method, in which a solid of revolution such as a disc, suspended by a wire, is allowed to oscillate about its axis in the liquid, against the restoring torque provided by the torsion in the wire, has been described in some detail by Stott (1933) who applied it to the investigation of the viscosity of molten tin. The chief disadvantage of this method is that contamination of the surface of the liquid, by oxide or otherwise, leads to a serious dragging effect on the rod connecting the disc to the suspension wire. It is difficult to eliminate this contamination in experiments on substances such as molten zinc. The method of enclosing the liquid inside a hollow sphere attached to a bifilar suspension and allowed to execute torsional oscillations has been developed and used with success by Andrade and Chiong (1936).

They discuss the calculation of the coefficient of viscosity from the observed damping of the oscillations and show that the method can be made to yield accurate results. Andrade and Rotherham (1936) give an alternative method of determining the viscosity by observing the forces necessary to maintain the oscillations at a constant amplitude.

The choice of material for the construction of the hollow sphere presents problems when it has to contain molten metal at high temperatures. The material must be immune to attack by the molten metal, and must be capable of being formed to the required shape without too much difficulty. Sintered alumina is a substance which withstands high temperatures and resists chemical attack, but hollow spheres which can be sealed with the metal inside are not easily made from this material. Accurately made cylindrical crucibles of sintered alumina are, however, readily available, and it seemed desirable that an investigation should be made of the possibility of determining the viscosity of molten metals by observing the oscillations of a cylindrical containing vessel left open at the top end. This method has been suggested by Knappwost (1948).

For a cylinder it is more difficult than in the case of a sphere to obtain a manageable expression from which the viscosity can be calculated in terms of the dimensions of the apparatus. It is possible, nevertheless, to use the method for absolute determinations. The theoretical considerations outlined below indicate how this may be done, and also provide the basis for a calibration formula, which can be determined by using liquids of known viscosity.

§ 2. THEORY OF THE METHOD

The oscillation of a closed cylinder of length short in comparison with its diameter has been considered by Meyer (1891), and viscosity determinations made on the basis of his analysis by Mutzel (1891).^{*} We shall consider here the case of a long cylinder open at one end. Let the cylinder, containing liquid, execute rotational oscillations about its axis (which we take to be the z -axis, in the vertical direction). Let (u, v, w) be the velocities of the liquid at the point (x, y, z) in the (x, y, z) directions, and let $p(x, y, z)$ be the pressure. Then, if (X, Y, Z) are the components of the body forces and ν the kinematic viscosity η/ρ where η is the coefficient of viscosity and ρ the density, the general equations of viscous flow are

$$\left. \begin{aligned} \frac{\partial u}{\partial t} + u \frac{\partial u}{\partial x} + v \frac{\partial u}{\partial y} + w \frac{\partial u}{\partial z} &= X - \frac{1}{\rho} \frac{\partial p}{\partial x} + \nu \nabla^2 u, \\ \frac{\partial v}{\partial t} + u \frac{\partial v}{\partial x} + v \frac{\partial v}{\partial y} + w \frac{\partial v}{\partial z} &= Y - \frac{1}{\rho} \frac{\partial p}{\partial y} + \nu \nabla^2 v, \\ \frac{\partial w}{\partial t} + u \frac{\partial w}{\partial x} + v \frac{\partial w}{\partial y} + w \frac{\partial w}{\partial z} &= Z - \frac{1}{\rho} \frac{\partial p}{\partial z} + \nu \nabla^2 w, \end{aligned} \right\} \dots\dots(1)$$

to which must be added the continuity equation

$$\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} + \frac{\partial w}{\partial z} = 0. \dots\dots(2)$$

Assuming that there is no motion in the z direction, that there are no body forces, and that the velocities are small enough to enable us to neglect the non-linear terms, the equations reduce to

$$\left. \begin{aligned} \frac{\partial u}{\partial t} &= -\frac{1}{\rho} \frac{\partial p}{\partial x} + \nu \nabla^2 u, \\ \frac{\partial v}{\partial t} &= -\frac{1}{\rho} \frac{\partial p}{\partial y} + \nu \nabla^2 v, \end{aligned} \right\} \dots\dots(3)$$

$$\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} = 0. \dots\dots(4)$$

Let us assume now that each particle of fluid moves in a circle with a velocity depending only on the distance r from the axis, on z and on t , and put

$$u = -\psi y, \quad v = \psi x,$$

where ψ is a function of r, z and t . The tangential velocity at a distance r is then $r\psi$ and the radial velocity zero; ψ is in fact the angular velocity of the liquid.

^{*} We are indebted to the referee for pointing out that work on the determination of viscosity by observing the oscillations of a closed cylinder containing liquid has been done by Okaya and Haegawa (1933) and Okaya (1936).

Multiplying the first of equations (3) by y , the second by x and subtracting, we have, in cylindrical polar coordinates, when there is no azimuthal variation of ψ

$$\frac{\partial^2 \psi}{\partial r^2} + \frac{3}{r} \frac{\partial \psi}{\partial r} + \frac{\partial^2 \psi}{\partial z^2} = \frac{1}{\nu} \frac{\partial \psi}{\partial t}. \quad \dots\dots(5)$$

We suppose that the system has settled down so that the liquid has the same period and decay as the containing vessel.

Then
$$\psi = \phi e^{\alpha t}$$

and if a is the radius of the cylinder, c the height of the liquid in the cylinder, and the origin is taken at the centre of the base of the cylinder,

$$\psi = \Phi e^{\alpha t} \text{ at } z=0 \text{ and at } r=a,$$

$$\frac{\partial \psi}{\partial z} = 0 \text{ at } z=c.$$

Equation (5) then becomes

$$\frac{\alpha}{\nu} \phi = \frac{\partial^2 \phi}{\partial r^2} + \frac{3}{r} \frac{\partial \phi}{\partial r} + \frac{\partial^2 \phi}{\partial z^2}, \quad \dots\dots(6)$$

with the boundary conditions

$$\left. \begin{aligned} \phi &= \Phi \text{ at } z=0 \text{ and at } r=a, \\ \frac{\partial \phi}{\partial z} &= 0 \text{ at } z=c. \end{aligned} \right\} \quad \dots\dots(7)$$

α is complex; in fact $\alpha = -\beta + i\gamma$, where $\beta = 2\delta/T$ and $\gamma = 2\pi/T$, δ being the logarithmic decrement and T the period of oscillation. The solution of (6) satisfying the boundary conditions (7) is

$$\phi = \frac{a\Phi J_1(mr)}{r J_1(ma)} + \frac{2\Phi}{r} \sum \frac{J_1(k_n r) m^2 \cosh l_n(z-c)}{J_0(k_n a) k_n (k_n^2 - m^2) \cosh l_n c}, \quad \dots\dots(8)$$

where $-m^2 = \alpha/\nu$ and J_0 and J_1 are Bessel functions of order 0 and 1 respectively. The summation is taken over the positive roots k_n of $J_1(k_n a) = 0$, l_n being given by $l_n^2 = m^2 - k_n^2$. The damping torque on the system due to the viscous drag of the liquid on the walls of the cylinder is

$$G = 2\pi\eta \left\{ a^3 \int_0^c \left(\frac{\partial \psi}{\partial r} \right)_{r=a} dz - \int_0^a \left(\frac{\partial \psi}{\partial z} \right)_{z=0} r^3 dr \right\}, \quad \dots\dots(9)$$

where
$$\frac{\partial \psi}{\partial r} = e^{\alpha t} \frac{\partial \phi}{\partial r} \quad \text{and} \quad \frac{\partial \psi}{\partial z} = e^{\alpha t} \frac{\partial \phi}{\partial z},$$

which, on evaluating with reference to (8), becomes

$$G = 2\pi\eta \left\{ -\frac{mca^2 J_2(ma)}{J_1(ma)} + 2m^4 a^2 \sum \frac{\tanh l_n c}{l_n k_n^2 (k_n^2 - m^2)} \right\} \Psi, \quad \dots\dots(10)$$

where $\Psi = \Phi e^{\alpha t}$ is the angular velocity of the cylinder containing the liquid. We can write $\Psi = d\theta/dt$ and $\theta = \Phi e^{\alpha t}/\alpha$ where θ is the angle through which the cylinder has turned from some zero position.

Equation (10) can be expressed as

$$G = L \frac{d\theta}{dt}, \quad \dots\dots(11)$$

and the equation of motion of the oscillating system, assuming that the external damping due to causes other than the viscous drag is small, is

$$I \frac{d^2\theta}{dt^2} + L \frac{d\theta}{dt} + f\theta = 0, \quad \dots\dots(12)$$

where I is the moment of inertia, and f the restoring torque per unit angle of rotation.

As is shown by Andrade and Chiong (1936), (12) may be written as follows on putting $L = M + iN$ and remembering that $\alpha = -\beta + i\gamma$.

$$\begin{aligned} I\alpha^2 + (M + iN)\alpha + f &= 0, \\ I(\beta^2 - \gamma^2) - \beta M - \gamma N + f &= 0, \\ -2I\beta\gamma - \beta N + \gamma M &= 0. \end{aligned}$$

i.e.

$$M = \beta \left\{ \frac{f}{\beta^2 + \gamma^2} + I \right\}, \quad N = \gamma \left\{ \frac{f}{\beta^2 + \gamma^2} + I \right\}.$$

Since $\beta = 2\delta/T$, $\gamma = 2\pi/T$, $f = 4\pi^2/T_0^2$ where T and T_0 are the periods of the full and empty cylinder respectively,

$$\left. \begin{aligned} M &= \frac{2I\delta}{T} \left(\frac{T^2}{T_0^2 \Delta^2 + 1} + 1 \right), \\ N &= \frac{2\pi I}{T} \left(\frac{T^2}{T_0^2 \Delta^2 + 1} - 1 \right), \end{aligned} \right\} \quad \dots\dots(13)$$

where $\Delta = \beta/\gamma = \delta/\pi$.

Since Δ is small, the expressions for M and N become

$$M = \frac{2I\delta}{T} \left(\frac{T^2}{T_0^2} + 1 \right); \quad N = \frac{2\pi I}{T} \left(\frac{T^2}{T_0^2} - 1 \right). \quad \dots\dots(14)$$

These equations contain only measurable quantities, so that M and N are known. Simple considerations show that if the external damping is small but not negligible, it is approximately accounted for by subtracting the logarithmic decrement observed when the cylinder is empty from that observed when the cylinder contains liquid, to obtain the value δ in the expression for M . The apparatus should, however, be designed to keep the external damping as low as possible.

The value of the coefficient of viscosity η in terms of the logarithmic decrement δ can theoretically be obtained by combining equations (10), (11) and (14) to give

$$\begin{aligned} \frac{2I\delta}{T} \left(\frac{T^2}{T_0^2} + 1 \right) &= \text{Real part of} \left[2\pi\eta \left\{ -mca^3 \frac{J_2(ma)}{J_1(ma)} \right. \right. \\ &\quad \left. \left. + 2\pi a^3 \sum_{n=1}^{\infty} \frac{\tanh L_n c}{L_n k_n^2 (k_n^2 - m^2)} \right\} \right] \quad \dots\dots(15) \end{aligned}$$

In order to use this, it is necessary to simplify the right-hand side. This we now proceed to do.

$$m^2 = -\frac{\alpha}{\nu} = \frac{1}{\nu} (\beta - i\gamma) = -\frac{i\gamma}{\nu} \left(1 + \frac{i\beta}{\gamma} \right).$$

Therefore

$$m = \left(\frac{\pi}{T\nu} \right)^{1/2} (1 - i) \left(1 + \frac{i\beta}{\gamma} \right)^{1/2}.$$

To a first approximation therefore

$$m = \left(\frac{\pi}{T\nu} \right)^{1/2} - i \left(\frac{\pi}{T\nu} \right)^{1/2}$$

and $ma = a \left(\frac{\pi}{T\nu} \right)^{1/2} (1-i) = \mu(1-i)$ say.

It is convenient to put $\zeta = \mu(1-i)$ and observe that

$$\frac{J_2(\zeta)}{J_1(\zeta)} = \frac{2}{\zeta} - \frac{J_0(\zeta)}{J_1(\zeta)}.$$

Since, in our experiments μ is large, we can use the well-known asymptotic expansions for $J_0(\zeta)$ and $J_1(\zeta)$, viz.

$$J_0(\zeta) \sim \left(\frac{2}{\pi\zeta} \right)^{1/2} \left\{ \left(1 - \frac{9}{128\zeta^2} \right) \cos \left(\zeta - \frac{\pi}{4} \right) + \frac{1}{8\zeta} \sin \left(\zeta - \frac{\pi}{4} \right) \right\},$$

$$J_1(\zeta) \sim \left(\frac{2}{\pi\zeta} \right)^{1/2} \left\{ \left(1 + \frac{15}{128\zeta^2} \right) \sin \left(\zeta - \frac{\pi}{4} \right) + \frac{3}{8\zeta} \cos \left(\zeta - \frac{\pi}{4} \right) \right\}.$$

Now $\cos(\zeta - \pi/4) = e^{\mu} e^{i(\mu - \pi/4)/2}$ and $\sin(\zeta - \pi/4) = e^{\mu} e^{i(\mu - \pi/4)/2} i$ on neglecting $e^{-\mu}$ compared with e^{μ} , so that

$$\begin{aligned} \frac{J_0(\zeta)}{J_1(\zeta)} &= \frac{1 - \frac{9}{128\zeta^2} + \frac{1}{8\zeta i}}{\frac{1}{i} \left(1 - \frac{15}{128\zeta^2} \right) + \frac{3}{8\zeta}} \\ &= \frac{1 + 8\zeta i}{8\zeta + 3i} \quad \text{neglecting terms in } \frac{1}{\zeta^2} \end{aligned}$$

which reduces to

$$\frac{2 + i(8\mu - 1)}{8\mu - 3} \quad \text{again neglecting terms in } \frac{1}{\mu^2} \text{ compared with unity.}$$

Hence

$$\zeta \frac{J_0(\zeta)}{J_1(\zeta)} = \mu \left(\frac{8\mu + 1}{8\mu - 3} \right) + i\mu$$

and

$$\begin{aligned} \zeta \frac{J_2(\zeta)}{J_1(\zeta)} &= 2 - \zeta \frac{J_0(\zeta)}{J_1(\zeta)} \\ &= 2 - \mu \left[\frac{8\mu + 1}{8\mu - 3} \right] - i\mu. \end{aligned}$$

Equation (15) may now be written

$$\frac{2L^2}{\pi} \left(\frac{T^2}{T_0^2} + 1 \right) = 2\pi\eta \left[-ca^2 \left\{ 2 - \mu \left(\frac{8\mu + 1}{8\mu - 3} \right) \right\} + K(\mu) \right], \quad (16)$$

where $K(\mu) = \text{Real part of } 2L^2 \sum \frac{\tanh k_n c}{k_n k_n^2 [(ak_n)^2 - \zeta^2]}$.

If the height of the liquid in the cylinder is not small compared with the diameter of the cylinder, $K(\mu)$ is much smaller than the other term on the right-hand side of (16). Physically, this means that most of the viscous drag takes place at the

curved surface of the cylinder as compared with the end. We can proceed now by using (16) to find an approximate value of η and of course μ , neglecting $K(\mu)$, and then using the approximate value of μ to evaluate $K(\mu)$ and so by a process of successive approximation arrive at a value of η satisfying the equation. It is found in practice that, after first obtaining an approximate value of μ , it does not involve much extra labour to compute $J_0(\zeta)/J_1(\zeta)$ without neglecting terms in $1/\zeta^2$, and thus to obtain numerically a more accurate expression than that given by the term containing $\{2 - \mu(8\mu + 1)/(8\mu - 3)\}$ in (16).

Calculation of $K(\mu)$ for various values of μ shows that it is very nearly a linear function of μ . It is worth tabulating this function if the apparatus is to be used unchanged for a large number of determinations.

An alternative procedure is to write (16) in the form

$$\delta \left(\frac{T^2}{T_0^2} + 1 \right) = A\eta T + B\mu\eta T + C \frac{\eta T}{\mu} \\ = A\eta T + B(\eta\rho T)^{1/2} + C \left[\frac{(\eta T)^2}{\rho} \right]^{1/2}$$

and to determine the constants A , B and C by observations on liquids of known viscosity. The cubic equation in $\eta^{1/2}$ can be solved by the usual methods.

§ 3. APPARATUS

The apparatus is described with reference to the sectional diagram of Figure 1.

The silica furnace tube K, 3 ft. 6 in. long and 5 in. internal diameter, is wound at E. The winding is arranged so as to ensure that a uniform temperature can be maintained over a length of the furnace safely in excess of that occupied by the sintered alumina crucible H which contains the molten metal under investigation. The dimensions of the crucible are: height 10 cm., internal diameter 3.9 cm., wall thickness 1 mm. Below the crucible are heat insulating bricks F which are supported by means of a long steel tube which rests on the lower steel endplate T of the furnace. The temperature of the furnace is controlled by a Cambridge regulator, and measured by means of a carefully calibrated platinum-platinum-rhodium thermocouple, used in conjunction with a potentiometer graduated in hundredths of a millivolt, but capable of being read by estimation to a tenth of this value. For the type of thermocouple used, one millivolt corresponds roughly to 100°C. The regulator is used to connect one or other of two alternative sources of current to the furnace winding. By careful adjustment of the relative magnitudes of the currents available from these two sources, and by maintaining the cold junction of the thermocouple at 0°C. in a vacuum flask containing melting ice, it is found that the temperature can be controlled to $\pm \frac{1}{2}^\circ\text{C}$. This variation takes place over a period which is long compared with that required to make a determination of the viscosity. The furnace tube is surrounded by the lagging J.

The crucible H is supported in a light steel cradle attached to a steel rod which passes through the steel tube supporting the bricks F in the upper half of the furnace. At its upper end the steel rod carries an aluminium disc D, having a graduated flange, which can be observed through a glass window in the brass cover G. The whole oscillating system consisting of the crucible, the connecting rod and aluminium disc is suspended by a phosphor bronze wire,

ten inches long and one hundredth of an inch in diameter, fastened to a brass rod L which fits closely into a brass sleeve, the joint being lubricated with vacuum grease. The suspended system is located vertically by the brass bush R clamped to the rod L. Slight rotation of R starts the oscillations.

The upper part of the apparatus, consisting of the brass cover G, the upper endplate T of the furnace, the upper insulating bricks F and the whole of the suspension system, can be removed as one integral part, the only vacuum seal to be broken being that between the silica tube and the furnace endplate. This seal is made with vacuum wax. The lower endplate is similarly sealed to the silica tube, so that this can be evacuated. The viscosity determinations are carried

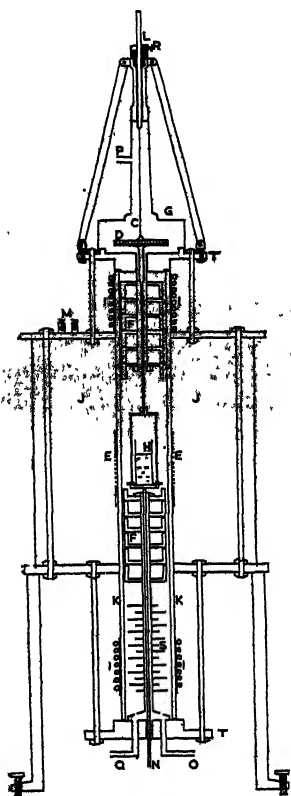


Figure 1.

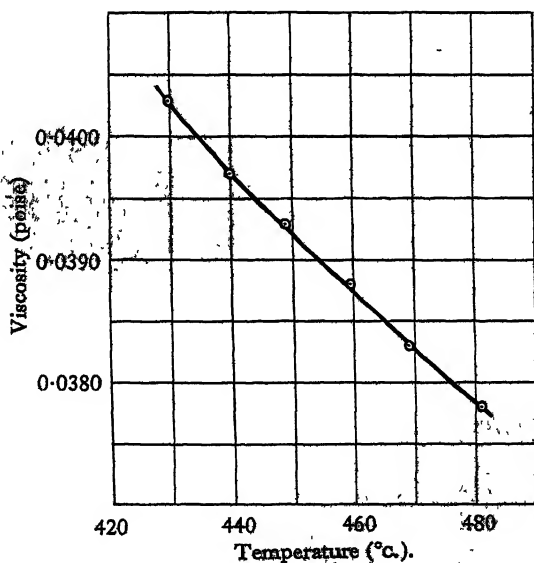


Figure 2.

out in an atmosphere of hydrogen, which after purification by passing through chromous chloride and potassium hydroxide, is dried and admitted through the side tube P in the brass cover G, and allowed to escape through the outlet tube Q at the bottom of the tube. Before the hydrogen is admitted the furnace tube is evacuated through the tube O in the bottom endplate.

In order to ensure that the phosphor bronze suspension wire be kept at a low constant temperature, and that the wax seals be kept cool, water cooling is provided at both ends of the silica tube.

Measurements of amplitudes and times of oscillation are made by observing the graduated aluminium disc D through a telescope placed a few feet away from the apparatus.

§ 4. EXPERIMENTAL PROCEDURE

A check determination was made of the viscosity of molten tin, the viscosity being calculated from the constants of the apparatus by the first of the methods described in § 2, that is without reference to observations on any other liquid. The value obtained was 0.01182 poise at 496.8° c. which agrees closely with the value 0.01187 obtained by Stott (1933).

In order to be able to use the second method given in § 2, the apparatus was calibrated, using liquids of known viscosity and density. The results are given in the next section. In all experiments the depth of liquid in the crucible was maintained the same, the amount of liquid necessary to ensure this being weighed beforehand.

The viscosity of molten zinc was then determined over a range of temperatures between 430° c. and 480° c. The material used was high purity zinc, for which we are indebted to The National Smelting Co., Ltd. The analysis gave the following quantities of impurities: Pb < 0.001% (probably 0.0005%), Cd 0.0001%, Cu 0.0001%, Fe 0.0005%, Ca a trace. The metal was melted in a separate furnace under a layer of wood charcoal, cooled and washed, first with 2% acetic acid solution to remove surface oxide and then with boiling distilled water. It was then placed in the viscometer and the temperature raised to 480° c. After 24 hours at this temperature, the logarithmic decrement was determined and further readings taken 12 hours and 24 hours later. Constant readings having been obtained corresponding to 480° c., the temperature was lowered to 470° c., and after an interval of 24 hours at this temperature the logarithmic decrement was again determined. Measurements were made, following the same procedure at 460°, 450°, 440° and 430° c. Trouble was at first experienced due to the vaporization of the zinc, but it was found that this could be eliminated by having a layer of powdered wood charcoal on the surface of the melt. Experiments on molten tin and molten zinc with and without such a layer of charcoal showed that the logarithmic decrement was not affected by its presence.

§ 5. RESULTS

(i) Calibration

The constants in the equation

$$\delta \left[\frac{T^2}{T_0^2} + 1 \right] = A\eta T + B(\eta\rho T)^{1/2} + C \left(\frac{(\eta T)^3}{\rho} \right)^{1/2}$$

were determined using five liquids: chloroform, water, carbon tetrachloride, benzene and tin, for which accurate values of η and ρ at various temperatures are available. The relevant data are given in Table 1. The value of δ given in the fourth column represents the observed value less the value of the logarithmic decrement obtained with the crucible empty at room temperature. This value was 0.003659 for the first four liquids and 0.001644 for the fifth. Each value given represents the mean of six separate observations.

The values of η and ρ for the first four liquids given in the table are obtained from *International Critical Tables*. The values for tin are taken from Stott's paper. Calculation of A , B and C from the above data by the method of least squares gives $A = -0.0362$, $B = 0.0438$, $C = 0.0306$. On using these values of A , B and C and the observed values of δ to calculate the viscosities η of the

quids from the calibration formula, the greatest discrepancy between d actual values is found to be 0.0002 poise. The values of δ , ρ and T of zinc (see below) are such that the determination of the viscosity ibration formula is an interpolation and not an extrapolation process.

(ii) The Viscosity of Zinc

e above values of A , B and C the viscosity of zinc was determined mperatures. The results are given in Table 2. The logarithmic i given in the table were obtained by subtracting from the observed garithmic decrement with the crucible empty at 455° c. This value

The values of the density were taken from a paper by Pascal and 914).

Table 1

Temperature (° c.)	Viscosity η (poise)	Density ρ (gm/cm ³)	Logarithmic decrement δ	Time of oscillation T (sec.)
23.00	0.005460	1.483347	0.007318	19.85
23.75	0.009216	0.99738	0.007286	19.86
19.75	0.009743	1.594539	0.009678	19.90
19.75	0.006495	0.879012	0.005692	19.87
496.8	0.01187	6.785	0.023824	20.09

Time of oscillation of empty crucible=19.81 seconds.

Table 2

Period (seconds)	Density (gm/cm ³)	Logarithmic decrement δ	Viscosity η (poise)
20.26	6.833	0.03929	0.0378
20.26	6.847	0.03955	0.0383
20.26	6.860	0.03981	0.0388
20.28	6.876	0.04008	0.0393
20.30	6.888	0.04027	0.0397
20.31	6.903	0.04056	0.0403

ion between the viscosity η and the temperature is shown in the ure 2.

(1934) has suggested a relationship between the viscosity of a liquid ture of the form

$$\eta v^{1/3} = Ae^{\theta/v}$$

he specific volume, θ the absolute temperature, and v and c are n fitting this relation to the results for molten zinc it was found that 09533 and $c = 81.266$ the calculated values of η agreed with the nes to within 0.1%.

(iii) Absolute Determinations

stants of the apparatus which enter into an absolute determination ty are the diameter of the crucible, the depth of the liquid in it, ment of inertia of the oscillating system. We have calculated the

viscosity of zinc using equation (16) and obtained the following values given in Table 3, the values of δ and ρ being those of Table 2.

Table 3

Temperature ($^{\circ}$ C.)	480.8	468.8	459.3	448.7	439.8	429.6
Viscosity (poise)	0.0380	0.0385	0.0390	0.0396	0.0400	0.0406

The greatest difference between these values and those of Table 2 amounts to 3 in the last decimal place, i.e. about three-quarters of one per cent.

(iv) Comparison with Previous Determinations

The only previous work on the viscosity of zinc of which we are aware is that of Gering and Sauerwald (1935), who determined the coefficients of viscosity of a number of metals by observing the motion of small quantities of the metal along a capillary tube under the influence of gas pressure. They obtained values for zinc in the temperature range 450° to 700° C. These values varied from 0.03168 poise at the lower to 0.01865 poise at the higher temperature. The results of the present paper are higher than this, e.g. our value at 448.7° C. is 0.0393 (Table 2), which may be compared with Gering and Sauerwald's 0.03168 at 450° C.

§ 6. CONCLUSION

The viscosity of pure zinc has been determined for various temperatures in the range 430° C to 480° C. by the oscillating cylinder method. It is intended to investigate the effect of adding small quantities of other metals such as aluminium to the zinc, and some preliminary measurements have already been carried out.

ACKNOWLEDGMENTS

The thanks of the authors are due to Sir Charles Goodeve, Director of The British Iron and Steel Research Association, who suggested the programme of which the work described in this paper forms part, for his continued interest, and to Mr. I. E. Thomas for valuable assistance in the construction of the apparatus.

REFERENCES

- ANDRADE, E. N. DA C., 1934, *Phil. Mag.*, **17**, 698.
 ANDRADE, E. N. DA C., and CHIONG, Y. S., 1936, *Proc. Phys. Soc.*, **48**, 247.
 ANDRADE, E. N. DA C., and ROTHERHAM, L., 1936, *Proc. Phys. Soc.*, **48**, 261.
 GERING, K., and SAUERWALD, F., 1935, *Z. anorg. allg. Chem.*, **223**, 204.
 KNAPPWOST, A. Z., 1948, *Metallkunde*, **39**, 314.
 MEYER, O. E., 1891, *Ann. Phys. Chem.*, **43**, 1.
 MÜTZEL, K., 1891, *Ann. Phys. Chem.*, **43**, 15.
 OKAYA, T., 1936, *Proc. Phys.-Math. Soc., Japan*, **18**, 268.
 OKAYA, T., and HASEGAWA, M., 1933, *Jap. J. Phys.*, **11**, 23.
 PASCAL, P., and JOUNIAUX, A., 1914, *Rev. Metall.*, **1**, 469.
 STOTT, V. H., 1933, *Proc. Phys. Soc.*, **45**, 530.

High Frequency Permeability of Ferromagnetic Materials

By R. MILLERSHIP* AND F. V. WEBSTER

Physics Department, University of Leeds

Communicated by R. Whiddington; MS. received 20th March 1950

ABSTRACT. Methods are described for the measurement of resistive and inductive permeabilities over a range of frequencies from 150 to 10,000 Mc/s. The measurements of the resistive permeabilities were carried out by comparing the attenuation of a coaxial transmission line with the ferromagnetic as the inner conductor with that of a coaxial line having a non-ferromagnetic as inner conductor. The inductive permeabilities were obtained from the wavelength in a coaxial line having the ferromagnetic as the inner conductor. The results show that at all frequencies the resistive permeability is greater than the inductive, and both decrease to unity at very high frequencies. The effective permeability, which is assumed to be complex, is deduced as a function of resistive and inductive permeabilities and plotted against frequency for a variety of materials. The experimental values are compared with theoretical values derived from two formal treatments.

§ 1. INTRODUCTION

THE process of magnetization of ferromagnetics under the influence of ultra-high-frequency fields has until recently received very little attention, probably because of the difficulty of producing an oscillatory field of sufficiently high frequency, but, as a result of extensive advances in ultra-high-frequency technique during the war, this difficulty has been largely removed. Unfortunately most of the previous investigations have been confined to very restricted frequency ranges; in fact, many have consisted of isolated determinations of permeability at a single frequency. Furthermore, the materials used in the past have often been of ill-defined composition. As the low field properties of magnetic materials are so strongly structure sensitive, it is only by using the same material (and preferably the same specimen) over the whole frequency range of interest that the character of the variation of permeability with frequency can be determined with any certainty.

A complete and critical account of the various methods which have been employed has been given by Allanson (1945). The methods fall into two distinct groups: those in which the permeability is determined from a measurement of the resistive losses of a circuit containing the ferromagnetic material, and those which are dependent upon the effective reactance of a similar circuit. The permeability deduced from measurements in the former category is usually denoted by μ_R , and that from measurements in the second category by μ_L . Of these two coefficients μ_R is always larger than μ_L , the difference between the two being covered formally by treating the permeability as a complex quantity. It is necessary to determine both μ_R and μ_L at the same frequency in order to calculate the complex permeability. This has not been generally realized, and, as a consequence, almost all the previous investigations have consisted of measurements of either μ_R or μ_L alone.

* Now with the National Coal Board.

In order to provide conditions under which the field may be introduced into the specimen it is convenient to use a transmission-line system incorporating the ferromagnetic material. The parameters of a transmission line which contains μ_R are the attenuation constant, the Q factor, and the characteristic impedance, whilst the wave velocity in the line gives the information necessary for the calculation of μ_L . Most of the measurements obtained by earlier workers were at frequencies of the order of 100 Mc/s., employing the open Lecher wire type of transmission line. Such a system would be unsuitable at higher frequencies as considerable power would be lost through radiation, and hand-capacity effects would also become significant. At these higher frequencies coaxial lines and waveguides have proved to be more satisfactory. One very accurate method would involve the use of a waveguide constructed partly of the ferromagnetic material: the disadvantages of this are that the magnetic properties of the material may be affected by machining, and that the waveguide could only be used over a limited range of frequencies.

One of the earlier theoretical treatments of magnetization at high frequencies was that developed by Arkadiew (1926) in which the permeability was introduced as a complex quantity. This theory contained much that is incompatible with the modern theory of ferromagnetism, and a theory based on more modern concepts was developed by Becker (1938, 1939), who assumed that the decrease in permeability was caused by a reduction of the field in the ferromagnetic due to microscopic eddy currents induced by domain boundary movements. Kittel (1946) attributed the reduction in the permeability to a reduced effective field across the domain boundary caused by the small penetration of the applied field due to the skin effect. Earlier Landau and Lifshitz (1935), considering the change of magnetization as due to the change in direction of the magnetic vector in the crystalline anisotropic field, predicted the occurrence of a resonance effect when the frequency of the applied field is equal to the Larmor frequency for gyroscopic rotation of the electron spins about the crystalline field direction. In metallic ferromagnetics, whether the magnetization change is due to translational boundary movements or to rotations of domain magnetization vectors, or both, the damping caused by microscopic eddy currents must be mainly responsible for the general decrease of permeability with frequency. These questions are further considered below (§ 4 (iv)).

§ 2. METHOD

(i) *Measurement of the Resistive Permeability*

The method used for the measurement of μ_R has been described in detail in a previous paper (Hodsman, Eichholz and Millership 1949) and hence only a general outline will now be given. A coaxial-line system is employed with the specimen as the centre conductor. This arrangement has the advantage that radiation can be propagated down the line in the fundamental mode over an infinite range of frequencies and so the introduction of the oscillatory field presents little difficulty. As the specimen is in the form of a wire, accurate machining to conform with the geometry of the system is eliminated. Furthermore, as the outer conductor is earthed, radiation is prevented from entering and leaving the system, which is consequently immune from external interference and instability due to hand-capacity effects.

The permeability is determined from the attenuation constant of the system, which is given by (see, for example, Jackson 1945, p. 50)

$$\alpha = \frac{9.95 \times 10^{-8} f^{1/2}}{\log(b/a)} \left[\frac{1}{a} \left(\frac{\mu_a}{\sigma_a} \right)^{1/2} + \frac{1}{b} \left(\frac{\mu_b}{\sigma_b} \right)^{1/2} \right] \text{ db/metre, } \dots (1)$$

where a , μ_a , and σ_a are the radius, permeability, and conductivity respectively of the inner conductor, b , μ_b and σ_b are those of the outer conductor, and f is the frequency of oscillation in megacycles per second. If a line of length l is used and the attenuation is measured with the inner conductor firstly of ferromagnetic wire and secondly of non-ferromagnetic wire, then the difference in attenuation is given by

$$\Delta(\alpha l) = (\alpha_1 - \alpha_2)l = \frac{9.95 \times 10^{-8} l f^{1/2}}{a \log(b/a)} \left[\left(\frac{\mu_1}{\sigma_1} \right)^{1/2} - \left(\frac{1}{\sigma_2} \right)^{1/2} \right] \text{ db. } \dots (2)$$

where μ_1 and σ_1 refer to the ferromagnetic specimen and σ_2 to the non-ferromagnetic.

In this way the effect of the outer conductor and of any other residual attenuation is eliminated. Measurements on the standing wave pattern in the line are transformed into values of input impedance for different terminations and plotted as circle diagrams of impedance, from which the attenuation constants are derived.

(ii) Measurement of the Inductive Permeability

It is apparent from the results of previous investigators that the dispersion in μ_L takes place at frequencies considerably below those at which the change in μ_R is most rapid. Preliminary investigations indicated that, at 1,500 Mc/s., μ_L was equal to unity for all the materials employed. Apparatus for the measurement of μ_L was therefore designed to cover the range from 150 to 1,500 Mc/s. only.

In order to measure μ_L under the same conditions as μ_R , a section of coaxial line was used in preference to the more easily constructed Lecher line, which ensured that the field configurations were precisely the same in both cases. The method employed is similar to that described by Glathart (1939), being based on the measurement of the wavelength in a coaxial line having the specimen as its centre conductor. Writing d_0 and d for the free space and measured half-wavelengths respectively the permeability is expressed as (Glathart 1939, p. 834)

$$\mu_L = 1.256 \left(\frac{a^2}{\rho} \right) \left(\log \frac{b}{a} \right)^2 \frac{(d_0^2 - d^2)}{d_0 d^4} \times 10^{13}, \dots (3)$$

where ρ is the resistivity of the inner conductor. Equation (3) may be solved for Δd giving

$$\Delta d = d_0 - d = d_0 \left[1 - \left\{ \frac{3.544 \times 10^6 a \log(b/a)}{(\mu \rho d_0)^{1/2} + 3.544 \times 10^6 a \log(b/a)} \right\}^2 \right] \dots (4)$$

By suitable choice of the dimensions of the system, the half-wave length shortening $d_0 - d$ may be made sufficiently large to allow an accurate determination of μ_L to be made; this is so when $a \rightarrow 0$ and $b \rightarrow \infty$. Having regard to mechanical stability, accurate centring of the inner conductor and convenient operating conditions, a fairly large diameter outer conductor containing a very fine inner conductor wire was designed. A schematic diagram of the apparatus is shown in Figure 1.

The wavelength is determined by altering the length of the line by means of a short-circuiting piston until two successive resonance peaks have been located. The distance moved by the piston between the resonance positions is then half the wavelength in the line. Resonance is observed by extracting a fraction of the power in the line by means of a small loop introduced through a slot at the input end of the outer conductor. The high-frequency current induced in the loop is rectified by a crystal and the output displayed directly on a meter. Figure 2 shows some typical resonance curves.

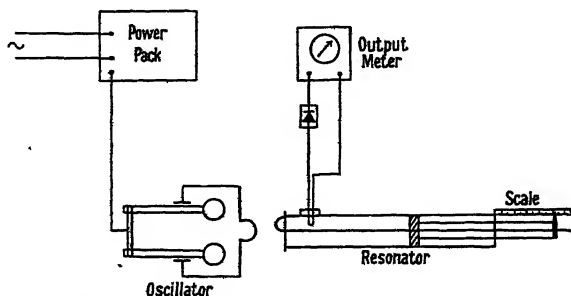


Figure 1. Layout of apparatus for measurement of μ_L .

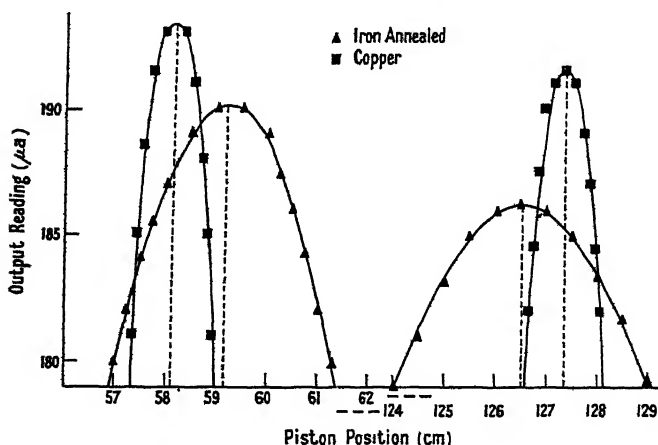


Figure 2. Typical resonance curves.

For the calculation of the free space half-wavelength equation (4) may be written

$$\Delta d_0 = d_0 \left[1 - \left(\frac{k}{(\rho d_0)^{1/2} + k} \right)^{1/2} \right], \quad \dots (5)$$

where Δd_0 is the difference between the free space half-wavelength and the measured half-wavelength for a non-magnetic inner conductor and $k = 3.544 \times 10^6 a \log(b/a)$. As Δd_0 is always small, d_0 in the above equation may be replaced by the measured half-wavelength.

The free space half-wavelength is then obtained as the sum of Δd_0 and the measured half-wavelength for a non-magnetic inner conductor, i.e.

$$d_0 = d \left[2 - \left(\frac{k}{(\rho d)^{1/2} + k} \right)^{1/2} \right]. \quad \dots (6)$$

§ 3. RESULTS

The following materials have been investigated, all in wire form: Swedish charcoal iron, white annealed and bright; 50/55 carbon steel, white annealed and bright; mumetal 6, hard drawn.

The percentage compositions and treatments of the specimens were as follows:

Iron: C, 0.03; Mn, 0.05; S, 0.002; Si, P, Cr, Ni, Co negligible.

Steel: C, 0.50-0.55; Mn, 0.5-0.6; S, 0.03; Si, 0.1; P, 0.02; Ni, 0.2; Cr nil.

Mumetal: Ni, 77; Fe, 15; Cu, 5; Cr, 2; Mn, 0.8; Si, 0.1.

The iron and steel specimens were hard drawn with 90% cold work. The white annealed specimens were annealed with a wet surface at a temperature of 700° c. in an inert atmosphere of town gas (CO, CO₂, N₂) and the bright specimens were left unannealed.

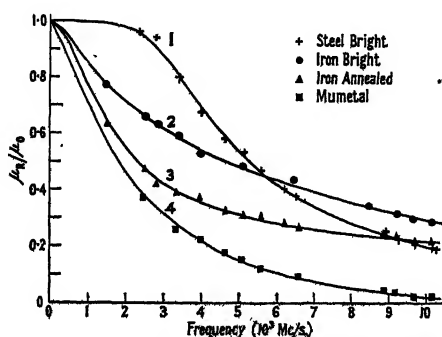


Figure 3. Relative reactive permeabilities.

Curve 1: Steel bright. Curve 2: Iron bright. Curve 3: Iron annealed. Curve 4: Mumetal.

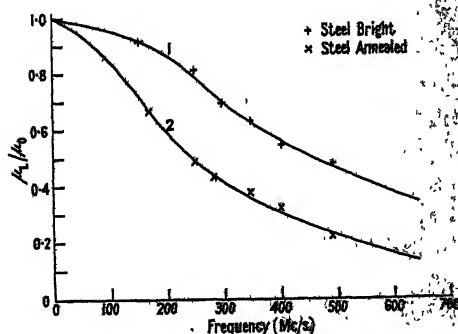


Figure 4. Relative inductive permeabilities.

Curve 1: Steel bright.
Curve 2: Steel annealed.

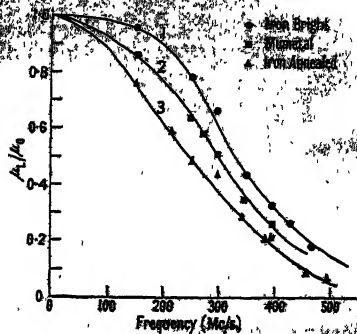


Figure 5. Relative inductive permeabilities.

Curve 1: Iron bright.
Curve 2: Mumetal.
Curve 3: Iron annealed.

In order to provide a firmer basis for comparison the results are shown as relative values, i.e. μ_r/μ_0 and μ_i/μ_0 in Figures 3, 4 and 5. The values of the initial permeability μ_0 are: iron annealed, 180; iron bright, 102; steel annealed, 103; steel bright, 65; mumetal, 440.

At any given frequency μ_r is greater than μ_i for all the materials employed. Both μ_r and μ_i appear to decrease to unity, although the frequency range

employed was not sufficiently great to obtain such a value for μ_R . No values below unity have been obtained, such as those reported by Birks (1948) for finely divided ferromagnetic compounds in paraffin wax. Owing to the lack of suitable oscillators it has not, as yet, been possible to obtain a sufficient number of readings to cover adequately the range where μ_L approaches unity.

§ 4. DISCUSSION OF RESULTS

(i) *Introduction*

An estimate of the strength of the high-frequency field that is applied to the specimen may be obtained from the equation for the power flowing down the line,

$$P = \pi \left(\frac{\mu_0}{\epsilon_0} \right)^{1/2} H_0^2 \log(b/a), \quad \dots\dots (7)$$

which is derived by integrating the time average Poynting vector over the cross section of the line. Assuming all the power produced by the oscillator is dissipated in the line an upper limit of 0.1 oersted is obtained for the field strength. The experimental values for the permeability therefore refer to magnetization in the reversible portion at the commencement of the magnetization cycle.

Any theory of the effects must explain the general decrease in the permeability, and also the difference which exists in the values obtained by resistive and inductive experimental methods.

(ii) *Effect of the Surface Condition*

It is apparent that the permeability for a high-frequency field is a property of the surface layer of the material and consequently will depend to some extent on the nature of the surface. It has in fact been suggested by Wien (1931), Michels (1931), and Procopiu and d'Albon (1937) that the general decrease in permeability with increasing frequency of the field is due to the existence of a thin surface layer (35–70 $\mu\mu$) of non-ferromagnetic material. Since the thinnest truly ferromagnetic films obtained by Sorenson (1924) and Tyndall (1924) are of the order of 20 $\mu\mu$, the existence of such a non-ferromagnetic layer appears to be unlikely. It is probable, however, that the domain structure in the surface layer will be different from that in the main body of the material: hence, if boundary movements are responsible for the magnetization at high frequencies, the magnetic properties will vary considerably for different skin depths and therefore for different frequencies.

Unfortunately, the permeability and conductivity always appear together as a ratio, i.e. μ/σ , in the equations for the attenuation and wavelength in a coaxial line. It is known that a change in conductivity takes place for non-ferromagnetics when the surface is etched or scratched and it is expected that a similar change takes place for ferromagnetics. If the surface is roughened the increased conducting path gives rise to a decreased conductance and permeance. These quantities are both dependent on the length of the conducting path and are reduced equally. Due to the reduction in the effective values of conductance and permeance the skin depth increases and brings about a corresponding increase in these quantities, the ratio μ/σ in the equations always remaining the same for different surface conditions. Thus it is very difficult to relate the permeability alone with surface condition, although it is intended to carry out experiments in this field in the future.

(iii) Difference between μ_R and μ_L

The most satisfactory approach to this problem was proposed by Arkadiew (1932) who treated the permeability as a complex quantity, i.e. of the form $\mu = \mu_1 - i\mu_2$. Considering a wire of ferromagnetic material, μ_R may be regarded as that value of μ which satisfies the theoretical equation for the resistance. Similarly μ_L emerges from the equation for the inductance.

The high-frequency resistance of an isolated conductor of radius r_0 per unit length is given by Jackson (1945, p. 38) as

$$R = \frac{1}{2\pi r_0} \left(\frac{\pi f \mu}{\sigma} \right)^{1/2}, \quad \dots\dots(8)$$

which gives (in C.G.S. units)

$$\frac{R}{R_0} = \frac{r_0}{2c} \left(\frac{2\pi\omega\mu_R}{\rho} \right)^{1/2}, \quad \dots\dots(9)$$

where R_0 is the D.C. resistance of the wire and ρ is the resistivity. Similarly

$$\frac{\omega L}{R_0} = \frac{r_0}{2c} \left(\frac{2\pi\omega\mu_L}{\rho} \right)^{1/2}. \quad \dots\dots(10)$$

Maxwell's equations for the specimen are

$$\left. \begin{aligned} -\frac{\partial H}{\partial r} &= \frac{4\pi E}{c\rho} \\ \frac{\partial E}{\partial r} &= i\frac{\omega\mu H}{c} \end{aligned} \right\}, \quad \dots\dots(11)$$

where r is the radial axis perpendicular to the direction of the wire.

Eliminating H ,

$$\frac{\partial^2 E}{\partial r^2} = i\omega \frac{4\pi\mu}{c^2\rho} E. \quad \dots\dots(12)$$

Substituting a complex permeability, $\mu = \mu_1 - i\mu_2$, gives

$$E = C \exp \left[\left\{ \frac{4\pi\omega}{c^2\rho} (\mu_2 + i\mu_1) \right\}^{1/2} r \right], \quad \dots\dots(13)$$

and

$$\frac{\partial E}{\partial r} = \left\{ \frac{4\pi\omega}{c^2\rho} (\mu_2 + i\mu_1) \right\}^{1/2} E. \quad \dots\dots(14)$$

The effective resistance R and inductance L are obtained from (Abraham and Becker 1932, p. 199)

$$\frac{R + i\omega L}{R_0} = i\frac{r_0}{2} \frac{4\pi\omega\mu}{c^2\rho} \left(\frac{E}{\partial E/\partial r} \right)_{r=r_0} \quad \dots\dots(15)$$

$$= \frac{r_0}{2} \left\{ \frac{4\pi\omega}{c^2\rho} (\mu_2 + i\mu_1) \right\}^{1/2} \quad \dots\dots(16)$$

$$= \frac{r_0}{2c} \left(\frac{2\pi\omega}{\rho} \right)^{1/2} \{ [(\mu_1^2 + \mu_2^2)^{1/2} + \mu_2]^{1/2} + i[(\mu_1^2 + \mu_2^2)^{1/2} - \mu_2]^{1/2} \} \quad \dots\dots(17)$$

and therefore

$$\left. \begin{aligned} \mu_R &= (\mu_1^2 + \mu_2^2)^{1/2} + \mu_2 \\ \mu_L &= (\mu_1^2 + \mu_2^2)^{1/2} - \mu_2 \end{aligned} \right\}, \quad \dots\dots(18)$$

from which

$$\left. \begin{aligned} \mu_1 &= (\mu_R \mu_L)^{1/2} \\ \mu_2 &= \frac{1}{2}(\mu_R - \mu_L) \end{aligned} \right\} \dots\dots (19)$$

These equations enable us to calculate μ_1 and μ_2 from the experimental values of μ_R and μ_L .

(iv) Theoretical Treatments

(a) Becker theory.

Becker (1938) considered the damping effect of microscopic eddy currents, associated with the movement of the domain walls, to be responsible for the general decrease in permeability with increasing frequency. As the domain wall moves, eddy currents, which are proportional to the velocity of movement, are induced in the surrounding medium. The eddy current field will be opposed to the change in magnetization and may thus be interpreted as a braking field which reduces the effective field acting on the wall. The displacements will therefore be reduced, resulting in a reduction of the permeability. This treatment leads to an expression of the form

$$\mu - 1 = \frac{\mu_0 - 1}{1 + i(\omega/\omega_1)}, \dots\dots (20)$$

where μ_0 is the D.C. initial permeability and ω_1 is a 'critical frequency' which, for a cubic domain of side l embedded in a ferromagnetic medium of susceptibility κ_0 , is expressed as

$$\omega_1 = \frac{3c^2}{4\pi^2 \sigma \kappa_0 l^2}. \dots\dots (21)$$

The high-frequency field is considered to be constant over the whole volume of the domain; owing to the small skin depth, however, the applied field may not even penetrate to the extent of the individual domains in the surface layer for centimetre waves. As the domain model assumed is over-simplified the equation for ω_1 is of doubtful validity, although equation (20) will still hold. Such an expression will result from any consideration of a damping force which is proportional to the velocity of movement of the boundary; and it is reasonable to suppose that these damping forces arise from the action of microscopic eddy currents.

Separating the real and imaginary parts of equation (20) we get

$$\left. \begin{aligned} \mu_1 - 1 &= \frac{\mu_0 - 1}{1 + (\omega/\omega_1)^2} \\ \mu_2 &= \frac{(\mu_0 - 1)(\omega/\omega_1)}{1 + (\omega/\omega_1)^2} \end{aligned} \right\}, \dots\dots (22)$$

the equations for the components of the complex permeability.

(b) Kittel theory.

A theory developed by Kittel (1946) relies on a more thorough consideration of the skin effect in a film one domain thick. The effective field acting on the domain boundary is the mean value of the applied field taken over the entire wall which moves as a rigid whole under its influence. The reduced permeability is thus accounted for in terms of the small effective field.

The theoretical curves are similar to those derived by Becker for frequencies beyond 3,000 Mc/s. but at lower frequencies they are very much steeper, and a

resonance peak appears on the initial part of the μ_R curve: of this there is no suggestion in the experimental curves. It seems therefore that the reduced penetration of the field due to the skin effect cannot be the sole cause of the ferromagnetic dispersion.

(c) *Damped resonance considerations.*

The results of experiments performed by Griffiths (1946) and Birks (1950), with a static field imposed on the specimen in addition to the transverse high-frequency field, indicate the existence of damped resonance phenomena. Resonance due to gyroscopic rotation of the electron spins about the field direction may occur when the frequency of the oscillating field is equal to the natural frequency of gyration given by the Larmor precession frequency, $\omega_0 = eH/4\pi mc$, where H is the strength of the field along the axis of rotation. With a superimposed static field the electron spins may be aligned throughout the material and a sharp resonance condition is obtained. A fuller treatment has been given by Kittel (1947, 1948) who showed that the resonance frequency depends on the shape of the specimen, and for a plane surface is given by $\omega_0 = \gamma(BH)^{1/2}$ (rather than $\omega_0 = \gamma H$), where $\gamma = e/4\pi mc = 2.80$ Mc/s/oersted.

Landau and Lifshitz (1935) were the first to consider resonance in the absence of an external biasing field, and deduced that it would occur about the anisotropy field direction. In an unmagnetized polycrystalline material the internal field will vary in direction throughout the volume and if strains are present there will also be a variation in magnitude; hence each domain will resonate at a different frequency and a broad flat resonance will be obtained instead of a sharp one.

In the experiments under discussion no static field was applied to the specimen and so the experimental results may be interpreted in terms of a damped resonance occurring about the crystal anisotropy field direction, the damping being caused by microscopic eddy currents.

The resonance dispersion equation for an oscillatory field transverse to the internal field is given by (Frenkel 1945)

$$\frac{\mu - 1}{\mu_0 - 1} = \frac{\omega_0^2}{\omega_0^2 - \omega^2 + 2i\omega\omega_1}, \quad \dots\dots(23)$$

where ω_0 is the 'resonance' frequency and ω_1 is the 'damping' frequency. Separating the real and imaginary parts of equation (23) we get

$$\left. \begin{aligned} \frac{\mu_1 - 1}{\mu_0 - 1} &= \frac{\omega_0^2(\omega_0^2 - \omega^2)}{(\omega_0^2 - \omega^2)^2 + 4\omega^2\omega_1^2} \\ \frac{\mu_2}{\mu_0 - 1} &= \frac{2\omega_0^2\omega\omega_1}{(\omega_0^2 - \omega^2)^2 + 4\omega^2\omega_1^2} \end{aligned} \right\}, \quad \dots\dots(24)$$

for the components of the complex permeability.

(d) *Summary of theories.*

The three methods of attack presented above may be summarized as follows:

(1) *Becker theory.* This theory is based entirely on damping produced by eddy currents on domain boundary movements. The formal treatment gives for μ the equation

$$\mu - 1 = \frac{\mu_0 - 1}{1 + i(\omega/\omega_1)}, \quad \dots\dots(25)$$

which contains one parameter that can be adjusted to give as close a fit to the experimental curves as possible.

(2) *Kittel theory*. Damping is not taken into account in this theory, the dispersion being wholly attributed to a reduction of the effective field acting on the domain wall because of the very small skin depth.

(3) *Damped resonance*. Rotations only are considered here and the formal treatment gives for μ

$$\frac{\mu - 1}{\mu_0 - 1} = \frac{\omega_0^2}{\omega_0^2 - \omega^2 + 2i\omega\omega_1}, \quad \dots\dots(23)$$

an equation with two parameters, both of which are adjusted to give the best fit to the experimental curves. It seems probable that these effects are present, not alone but all together in varying degrees, and, therefore, it is not to be expected that a perfect fit to the experimental curves will be obtained by considering any one effect. The aim is to obtain a general fit which may give some indication of the type of mechanism responsible for the magnetization at high frequencies, e.g. rotational movements of the spins or translational movements of the domain walls.

(v) *Comparison of the Experimental Results with Theories*

(a) *Becker theory*.

A comparison of the experimental results for the complex permeability with values of μ_1 and μ_2 calculated from equation (22) are given in Figures 6, 7 and 8.

The theoretical curves are fitted to the experimental results by making ω_1 (in equation (22)) approximately equal to the frequency at which the experimental value of μ_2 reaches a maximum. From the values of ω_1 the sizes of the domains are calculated using equation (21). These are: iron annealed $l = 1.7 \times 10^{-4}$ cm.; steel annealed $l = 1.4 \times 10^{-4}$ cm.; mumetal $l = 1.6 \times 10^{-4}$ cm. These values are of the right order of magnitude although little reliance can be placed on them because of the over-simplification of the domain model. It has been pointed out by Landau and Lifshitz (1935) and Néel (1944) that the domains in the body of the material will most likely be in the form of plane parallel sheets, whereas the surface layer will consist of small domains of triangular cross section which provide 'closures of flux' for the main internal domains and thus prevent the appearance of free poles at the surface.

Although the general forms of the experimental and theoretical curves are similar, the frequency at which μ_1 is equal to μ_2 is in all cases lower than that at which μ_2 is a maximum, whereas the theory shows that the μ_1 curve should pass through the maximum of the μ_2 curve. Also, the real part of the permeability decreases more rapidly than the theory indicates; this was observed by Johnson and Rado (1949) who experimented at frequencies of 200 and 975 Mc/s. with a superimposed static field applied to the specimen. The theory gives a much larger and sharper maximum for μ_2 than the experimental curve suggests.

(b) *Damped resonance*.

Theoretical values of the complex permeability are derived from equation (24) and are plotted against frequency together with experimental values in Figures 9, 10 and 11.

The curves are fitted to the experimental values by choosing ω_0 and ω_1 to produce a μ_2 curve which has a maximum approximating to the experimental maximum. The condition for a maximum in μ_2 is

$$3\omega^4 - 2\omega^2(\omega_0^2 - 2\omega_1^2) - \omega_0^4 = 0. \quad \dots\dots (25)$$

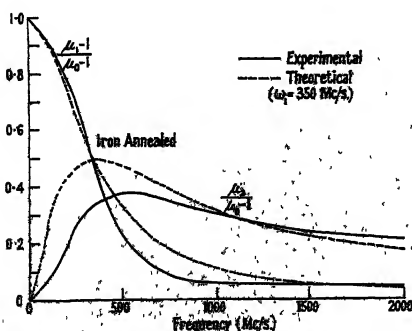


Figure 6. Comparison of experimental results with the Becker theory—Iron annealed.

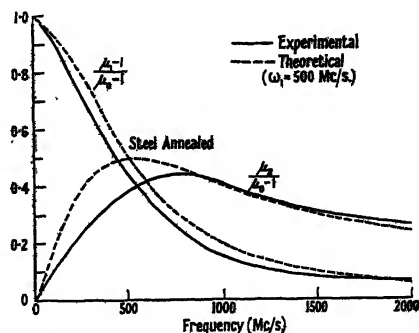


Figure 7. Comparison of experimental results with the Becker theory—Steel annealed.

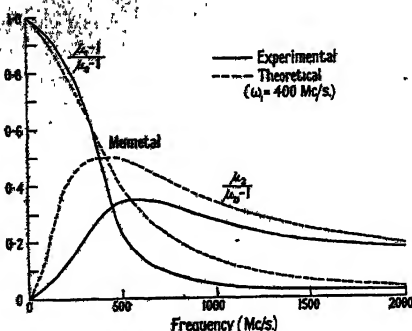


Figure 8. Comparison of experimental results with the Becker theory—Mumetal.

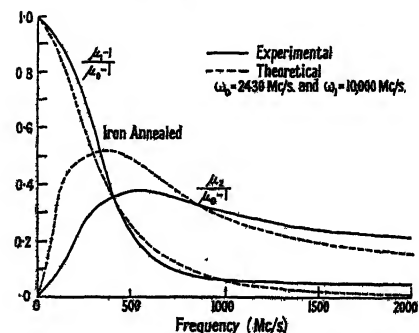


Figure 9. Comparison of experimental results with the damped resonance theory—Iron annealed.

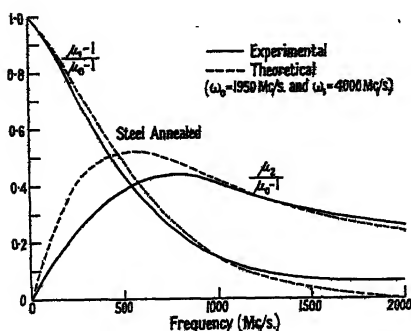


Figure 10. Comparison of experimental results with the damped resonance theory—Steel annealed.

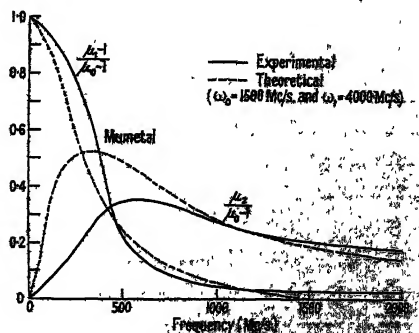


Figure 11. Comparison of experimental results with the damped resonance theory—Mumetal.

Due to a lack of suitable oscillators covering the range of frequencies between 500 and 1,000 Mc/s. the exact location of the maximum on the experimental curve for μ_2 cannot be obtained without further results for μ_2 in this range.

In some respects the type of curve obtained from considerations of a damped resonance expression approximates more closely to the experimental results than that drawn from an expression of the Becker type. The μ_1 and μ_2 theoretical curves intersect at a frequency slightly lower than that at which μ_2 is a maximum, which is in better agreement with experiment. Also, the μ_1 curve is nearer to the experimental results than the Becker μ_1 curve.

There are certain points, however, in which the experimental and theoretical curves do not agree. Firstly, the theory indicates that μ_1 becomes unity at a frequency equal to ω_0 and becomes less than unity at frequencies greater than ω_0 . Secondly, the theoretical curve (like the Becker curve) has a much larger and sharper maximum than the experimental points suggest.

A comparison of the resonance frequency with that which would be obtained for a resonance of the spins about the anisotropy field may be made by substituting the value of the field in the Larmor expression. The anisotropy field is given (Birks 1950) by $H = 2K/I_0$, where K is the anisotropy constant and I_0 is the saturation magnetization. For iron, this is approximately 500 oersteds, which would give a Larmor frequency of 1,400 Mc/s. The resonance frequency for iron obtained from the experimental curves is therefore of the same order of magnitude as the Larmor frequency for the electron spins about the anisotropy field. The Larmor frequencies for steel and mumetal are approximately 1,400 and 120 Mc/s. respectively, which, especially in the case of mumetal, is not in very good agreement with the value of ω_0 substituted in the equation for μ . It is unlikely that an entirely different mechanism is operative for mumetal, and seems to indicate that more than one process is involved, e.g. a combination of boundary movements and spin rotations, both being damped by microscopic eddy currents.

§ 5. CONCLUSION

Experimental values of the permeability have been obtained from both resistive and inductive measurements on a circuit containing the ferromagnetic. The values of μ_R and μ_L both decrease to unity as the frequency of the applied field is increased and μ_R is always greater than μ_L , the difference between them being covered formally by supposing the permeability to be complex.

The experimental values of the components of the complex permeability are compared with values obtained from two theoretical treatments, one based on boundary movements and the other on rotations; a fairly good fit is obtained in each case. The resonance frequency for iron calculated from damped resonance considerations approximates to the Larmor precession frequency of the electron spins about the anisotropy field, but the divergence for mumetal suggests the existence of more than one mechanism. It seems probable that the magnetic dispersion is due to eddy current damping of translational and rotational movements to approximately the same degree, the situation being further complicated by the possibility that skin depth effects of the type suggested by Kittel are responsible in some measure for the dispersion.

ACKNOWLEDGMENTS

We are greatly indebted to Professor R. Whiddington, under whose direction the work has been carried out, and to Professor E. C. Stoner for his continued interest and valuable advice. We wish to thank the Captain Superintendent, Admiralty Signal Establishment, for the generous loan of equipment, Messrs. John

Rigby & Sons Ltd., who provided the iron and steel specimens, Messrs. Telephone Construction & Maintenance Co. Ltd., who provided the mumetal wire and Mr. K. S. Driver for help with the proofs. One of us (F.V.W.) wishes to express his indebtedness to the Department of Scientific and Industrial Research for a grant.

REFERENCES

- ABRAHAM, M., and BECKER, R., 1932, *The Classical Theory of Electricity and Magnetism* (London: Blackie).
 ALLANSON, J., 1945, *J. Instn. Elect. Engrs.*, **92**, Pt. III, 247.
 ARKADIEW, W., 1926, *Ann. Phys., Lpz.*, **81**, 649; 1932, *Z. Phys.*, **74**, 5-6, 396.
 BECKER, R., 1938, *Phys. Z.*, **39**, 856; 1939, *Ann. Phys., Lpz.*, **36**, 340.
 BIRKS, J. B., 1948, *Proc. Phys. Soc.*, **60**, 282; 1950, *Proc. Phys. Soc. B*, **63**, 65.
 FRENKEL, J., 1945, *J. Phys., U.S.S.R.*, **9**, 299.
 GLATHART, J. L., 1939, *Phys. Rev.*, **55**, 833.
 GRIFFITHS, J. H. E., 1946, *Nature, Lond.*, **158**, 670.
 HODSMAN, G. F., EICHOLZ, G., and MILLERSHIP, R., 1949, *Proc. Phys. Soc. B*, **62**, 377.
 JACKSON, W., 1945, *High Frequency Transmission Lines* (London: Methuen).
 JOHNSON, M. H., and RADO, G. T., 1949, *Phys. Rev.*, **75**, 841.
 KITTEL, C., 1946, *Phys. Rev.*, **70**, 281; 1947, *Ibid.*, **71**, 270; 1948, *Ibid.*, **73**, 155.
 LANDAU, L., and LIFSHITZ, E., 1935, *Phys. Z. Sowjet*, **8**, 153.
 MICHELS, R., 1931, *Ann. Phys., Lpz.*, **8**, 877.
 NIEL, L., 1944, *J. Phys. Radium*, **5**, 241.
 PROCOPIU, S., and D'ALBON, G., 1937, *C. R. Acad. Sci., Paris*, **205**, 1373.
 SORENSEN, A. J., 1924, *Phys. Rev.*, **24**, 658.
 TYNDALL, E. P. T., 1924, *Phys. Rev.*, **24**, 439.
 WIEN, M., 1937, *Ann. Phys., Lpz.*, **8**, 899.

The Production of Pulsed Magnetic Fields using Condenser Energy Storage

By K. S. W. CHAMPION

Electron Physics Department, University of Birmingham

Communicated by J. Sayers; MS. received 21st March 1950

ABSTRACT. An economical way of producing a pulsed magnetic field is to use condensers as the energy source and an air-cored magnet coil which, with the storage condensers, forms an oscillatory circuit. By means of a switch the coil is connected to the charged condensers and then disconnected precisely at the end of one cycle. In this way all the stored energy is used to produce the field, while most of it is recovered at the end of the pulse, the only appreciable losses being due to the resistance of the coil and switch.

The design of coils for use in the way described above is considered in detail. Expressions and curves are presented which enable the field produced by a specified coil and condenser combination to be readily calculated. Conversely, the data can be used to design the most satisfactory coil for any practical application. An example is given of the design and construction of an air-cooled coil.

§1. INTRODUCTION

TO investigate certain fundamental properties of low pressure arcs it became necessary to study the effect on them of strong magnetic fields. As the arcs in question were to be pulsed and the field required was inconveniently large for a conventional iron electromagnet, it was decided to pulse the magnetic field. The most satisfactory sources of the current pulse for such fields are

suitable generators or condensers. The design of coils for use with generators has been considered very fully by Cockcroft (1928).

However, very often condensers are a more satisfactory source of energy when such considerations as maintenance are taken into account, and, in the following, the design of suitable coils for use with a condenser energy source is considered. It is hoped that these data will be of use when it is necessary to design efficient coils for producing strong pulsed magnetic fields, for example, those for use with cloud chambers.

§ 2. THEORY OF PULSED MAGNETIC FIELD

If it is not necessary that the magnetic field pulse be square, a convenient and economical way of producing it is as follows. With the circuit as in Figure 1 the switch is closed when a pulse is desired and opened again precisely at the end of one cycle. By this means all the stored energy is used to produce the magnetic field, while most of it is returned to the condenser at the end of the cycle, the only appreciable losses being due to the resistance of the coil and switch. In practice it is usually found that an electronic switch is much more satisfactory than a mechanical switch.

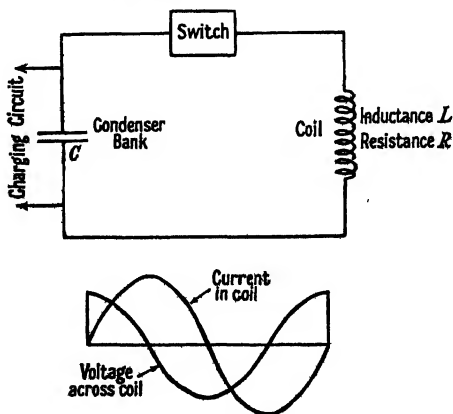


Figure 1. If R is not negligible the phase of the current is shifted and the waves are damped.

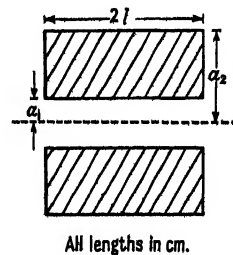


Figure 2.

If it were possible to convert all the energy stored in the condenser into a uniform magnetic field H_0 within the cylinder enclosed by the inner diameter of the coil, then the following relation would hold

$$\frac{H_0^2}{8\pi} \pi a_1^2 2l = \frac{1}{2} CV^2 \cdot 10$$

$$\text{i.e. } H_0 = \left(\frac{20C}{l} \right)^{1/2} \frac{V}{a_1} \quad \dots\dots(1)$$

where H_0 is in gauss, C in μF . and V in volts and it is assumed that the permeability is unity. The symbols representing the coil dimensions are defined in Figure 2.

Thus the actual magnetic field H at the centre of a coil can be expressed by

$$H = \left(\frac{20C}{l} \right)^{1/2} \frac{V}{a_1} K \quad \dots\dots(2)$$

where K is a positive quantity less than one. It is convenient to split K into the product of two quantities S and J where S depends only on the shape of the

coil and arises from the fact that the field is not uniform and not all contained in the cylinder enclosed by the inner diameter of the coil. J , on the other hand, is a factor to correct for the reduction in the first current peak (and, hence, the maximum magnetic field) due to the finite resistance of the coil. The values of S and J each lie between zero and one.

$$H = \left(\frac{20C}{l} \right)^{1/2} \frac{V}{a_1} SJ. \quad \dots\dots(3)$$

The value of the magnetic field at the centre of a coil may be expressed in the form

$$H = \frac{2\pi nI}{10(a_2 - a_1)} \cdot \ln \frac{a_2 + (l^2 + a_2^2)^{1/2}}{a_1 + (l^2 + a_1^2)^{1/2}} \quad \dots\dots(4)$$

where n is the number of turns in the coil (assumed to be uniformly distributed)

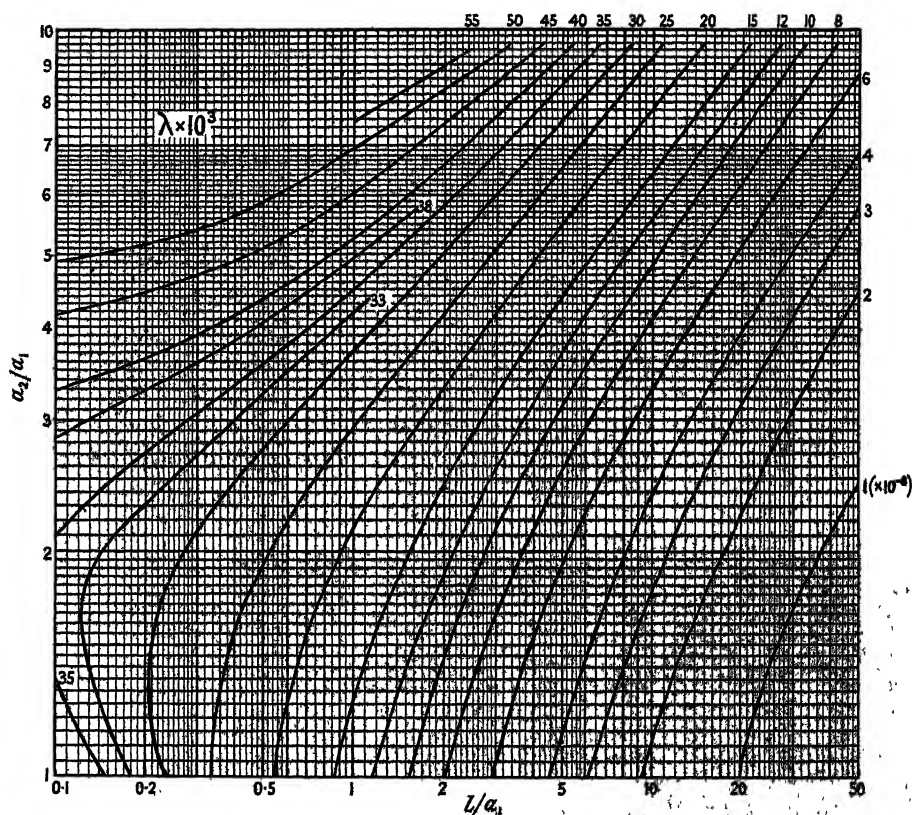


Figure 3. $L = n^2 a_1 \lambda$ or $n = (L/a_1 \lambda)^{1/2}$. L in μH , a_1 in cm .

and I is the current in amperes. Now $n = (L/a_1 \lambda)^{1/2}$, as indicated by the curves for inductance* (Figure 3), and the maximum value of I is

$$I = \frac{V_0 - R I_0}{(R^2 + L^2 \omega^2)^{1/2}},$$

where $\omega^2 = \frac{1}{LC} - \frac{R^2}{4L^2}$ and $T = \frac{2\pi}{\omega}$.

* These curves were calculated from data in a paper by Grover 1922.

Now

$$\frac{I}{V} \left(\frac{L}{C} \right)^{1/2} = \frac{e^{-RT/8L} \cdot \left(\frac{L}{C} \right)^{1/2}}{\left(R^2 + \frac{L^2 4\pi^2}{T^2} \right)^{1/2}}$$

$$= \frac{e^{-RT/8L} \left(1 + \frac{R^2 T^2}{16\pi^2 L^2} \right)^{1/2}}{\left(1 + \frac{R^2 T^2}{4\pi^2 L^2} \right)^{1/2}}$$

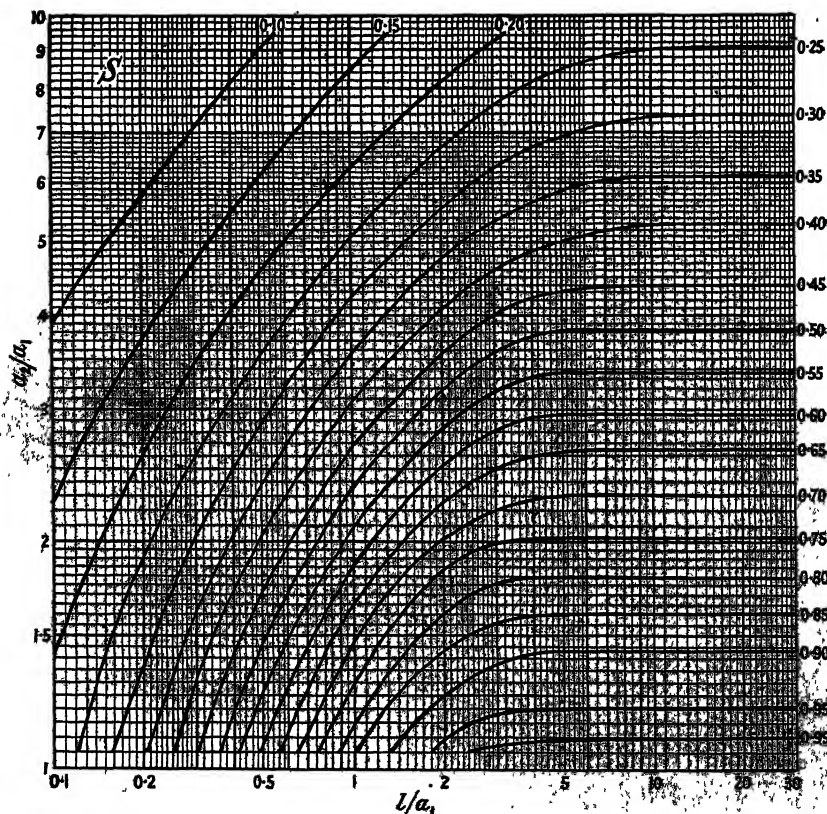


Figure 4.

$$H = \left(\frac{20C}{l} \right)^{1/2} \frac{V}{a_1} SJ \quad H \text{ in gauss, } C \text{ in } \mu\text{F, } l \text{ and } a_1 \text{ in cm, } V \text{ in volts.}$$

This quantity, related to the first peak value I of the current, depending only on RT/L and such that its value is unity when R is zero, is the quantity defined above.

$$J = \frac{I}{V} \left(\frac{L}{C} \right)^{1/2} = \frac{e^{-RT/8L} \left(1 + \frac{R^2 T^2}{16\pi^2 L^2} \right)^{1/2}}{\left(1 + \frac{R^2 T^2}{4\pi^2 L^2} \right)^{1/2}} \quad \dots\dots(5)$$

If we substitute for n and I in (4) we obtain

$$H = \frac{2\pi L^{1/2} V J C^{1/2}}{10(a_2 - a_1) a_1^{1/2} \lambda^{1/2} L^{1/2}} \ln \frac{a_2 + (l^2 + a_2^2)^{1/2}}{a_1 + (l^2 + a_1^2)^{1/2}}.$$

From comparison of this expression and (3)

$$S = \frac{\pi \left(\frac{l}{a_1}\right)^{1/2}}{10 \times 5^{1/2} \left(\frac{a_2}{a_1} - 1\right) \lambda^{1/2}} \ln \frac{\frac{a_2}{a_1} + \left(\frac{l^2}{a_1^2} + \frac{a_2^2}{a_1^2}\right)^{1/2}}{1 + \left(\frac{l^2}{a_1^2} + 1\right)^{1/2}}. \quad \dots\dots (6)$$

This is the required shape factor since it is a function of l/a_1 and a_2/a_1 only. Curves of S are plotted in Figure 4. As would be expected, the value of S tends to unity for long, thin coils.

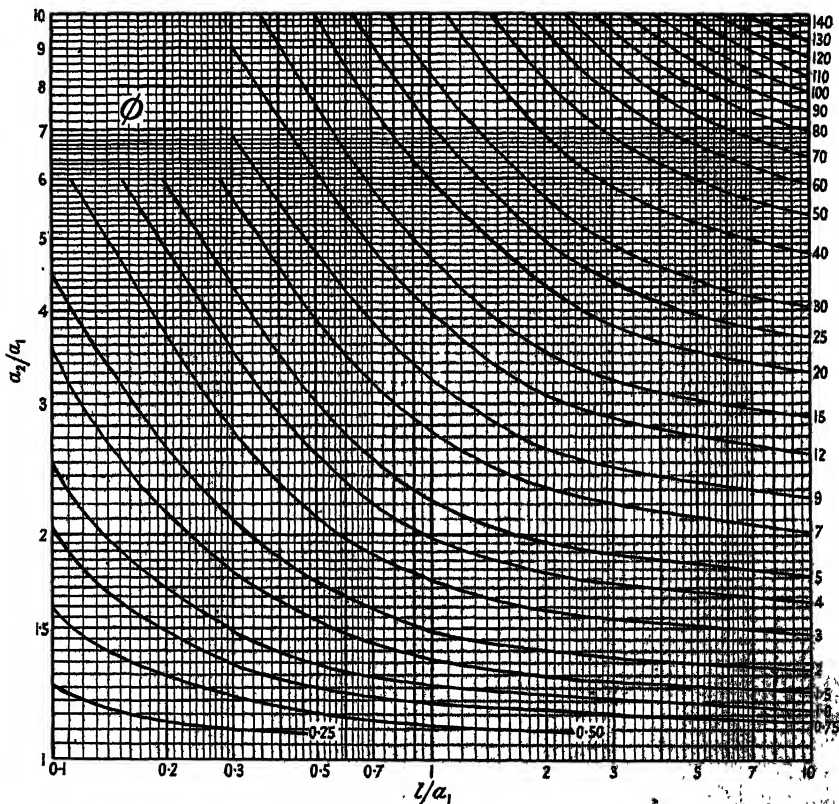


Figure 5.

$R = \frac{\rho 10^9}{a_1^2 \lambda_s \phi}$ R resistance of coil in ohms., ρ specific resistance (ohm. cm.),
 $L = \frac{4\pi \times 10^{-9}}{9} \frac{L^2}{a_1} \phi$ L inductance of coil in henries, a_1 internal coil radius (cm.), λ_s space factor.

Now $R/L = 10^9 \rho / a_1^2 \lambda_s \phi$ where ρ is the specific resistance of the material of the coil (in ohm.cm.), λ_s the space factor* and ϕ a quantity which is a function only of the shape of the coil.

$$\phi = \frac{2}{\pi} \times 10^3 \frac{l}{a_1} \lambda \left(\frac{a_2/a_1 - 1}{a_2/a_1 + 1} \right)$$

* The proportion of the coil winding volume occupied by conducting material is called the space factor.

Values of $\phi(\dagger)$ are plotted in Figure 5 and values of J as a function of RT/L are set out in the Table. For values of RT/L up to 1.8, J may be represented by $(1 - RT/8L)$ with an accuracy closer than 0.2%.

Calculation of K for a Coil

Thus, to calculate the magnetic field produced at the centre of a coil, operated in the manner previously described, it is necessary to calculate K . This can be done by making use of the curves in Figures 4 and 5 and the values in the Table. The value of K is completely determined by l , a_1 , a_2 , λ_s , ρ and T .

As an example, consider a coil with $l=5$ cm., $a_1=2.5$ cm., $a_2=4$ cm., $\lambda_s=0.7$, $\rho=1.724 \times 10^{-6}$ ohm.cm. and $T=1$ millisecc. In this case $l/a_1=2$ and $a_2/a_1=1.6$ and thus $\phi=3.15$ and $R/L=125.1$. Now $RT/L=0.1251$ and hence $J=0.984$; since $S=0.803$ the value of K is 0.790.

Table

RT/L	J	Difference per 0.1 of RT/L	RT/L	J	Difference per 0.1 of RT/L
0	1		2.5	0.609	0.0113
0.01	0.999		3	0.638	0.011
0.05	0.994		3.5	0.586	0.0104
0.1	0.987		4	0.537	0.0098
0.2	0.975				Per 1 of RT/L
0.3	0.962		5	0.451	0.086
0.4	0.950		6	0.379	0.072
0.5	0.937	0.0125	7	0.319	0.060
0.6	0.925		8	0.270	0.049
0.8	0.900		10	0.195	0.038
1.0	0.875		12	0.143	0.026
1.2	0.850		15	0.0923	0.0169
1.4	0.825		20	0.0463	0.0092
1.6	0.800		25	0.0239	0.0045
1.8	0.775		30	0.0125	0.0023
2.0	0.751	0.012	40	0.0035	0.0009
2.2	0.727	0.012	50	0.00099	
			∞	0	

If the coil is used with a condenser bank containing 1,000 joules a field of 19,980 (approximately 20,000) gauss will be produced at its centre. Note that, with the above specifications, the ratio R/L is determined. Also, since

$$T = 2\pi \left(\frac{1}{LC} - \frac{R^2}{4L^2} \right)^{-1/2}$$

the product LC is determined. Hence, when any one of R , L , C is specified, values of the other two are fixed.

Limitations of the Data

Before considering the design of coils it is desirable to point out the limitations of the application of the data contained in this paper. In the curves for inductance (Figure 3) it has been assumed that the inductance is independent of the space factor, which is true when the latter is near unity. According to Grover (1922, p. 462), "only when the wires are relatively far apart will the correction (to be

\dagger The quantity ϕ is identical with Cockcroft's ϕ_1 (1928). However, additional values have been calculated for the curves in Figure 5.

added to the inductance) exceed a few tenths of 1%". As these inductance curves have been used in the calculation of the ϕ and S curves any deviation from the former will be reflected in the latter. Secondly, the data only apply accurately to relatively low frequencies, as at radio frequencies the calculated inductance is reduced slightly and the resistance increased by the skin effect. Also, at these frequencies the coil capacity becomes appreciable and radiation losses may no longer be negligible.

§3. EXAMPLE OF THE DESIGN OF A COIL

As a source of energy a condenser of capacity $7\mu\text{F.}$ and maximum working potential 25 kv. is available. Consider the design of a coil to produce a maximum magnetic field with this source, given that its length is to be 20 cm. and internal radius 2.9 cm. The pulse length is to be 10 milliseconds and, from considerations of the insulation requirements, a space factor* of 0.7 seems possible. Assume that the mean temperature of the coil is 20°C. , then the value of ρ is the specific resistance of copper at this temperature. Values of K calculated from these data are plotted as a function of a_2/a_1 in Figure 7. K has a maximum value of 0.759 when $a_2/a_1 = 1.5$.

Heating of Coil

At this point it is necessary to consider the energy dissipated in the coil and the resultant temperature rise per pulse. The temperature rise per pulse

$$\delta t = \int_0^T \frac{Ri^2 dT}{v\sigma d}$$

where i is the current at any instant, v the volume of the coil windings and d and σ respectively the density and specific heat of the material of the coil. Assume that R is constant during a pulse and take $\frac{1}{2}I_1^2$ as the mean value of i^2 where I_1 is the amplitude of the current vector at $\frac{1}{2}T$. Thus, with sufficient accuracy

$$\delta t = \frac{RI_1^2 T}{2v\sigma d}$$

where $I_1 = VJ_1(C/L)^{1/2}$ with $J_1 = 1 - RT/4L$ for small values of RT/L . Values of δt calculated from this expression are plotted in Figure 7.

Now it is desired to operate this coil with one pulse per second for considerable periods. With $a_2/a_1 = 1.5$ and neglecting thermal conduction the rise in temperature per minute of operation would be 32.7°C. Even with mica insulation and a Tufnol former it is not advisable to have an increase in temperature greater than 150°C. and for this reason, a larger coil is required. For example, with $a_2/a_1 = 2.4$ the temperature rise is only 4.1°C. per minute. Such a coil, even with no loss of heat, could be operated for at least 30 minutes and, with efficient cooling of the outer surface, this time could be extended indefinitely.

Choosing a coil of this size the appropriate wire gauge is now to be determined. The resistance per unit length

$$\frac{Ra_1^{1/2}\lambda^{1/2}}{\pi L^{1/2}(a_1 + a_2)} = 1.074 \times 10^{-4} \text{ ohm/cm.}$$

* A higher space factor could be achieved if wire of rectangular section were used.

Hence 17 s.w.g. is the nearest gauge as it has a resistance of 1.068×10^{-4} ohm/cm. If wire of the calculated resistance were used 3,520 turns would be required.

A coil based on this design has been constructed and is illustrated in Figure 6. There are only 3,410 turns on the coil as its length is slightly less than



Figure 6. Coil in its mounting. The coil can be supported with its axis either vertical or horizontal by screwing the base plate to the appropriate part of the frame.

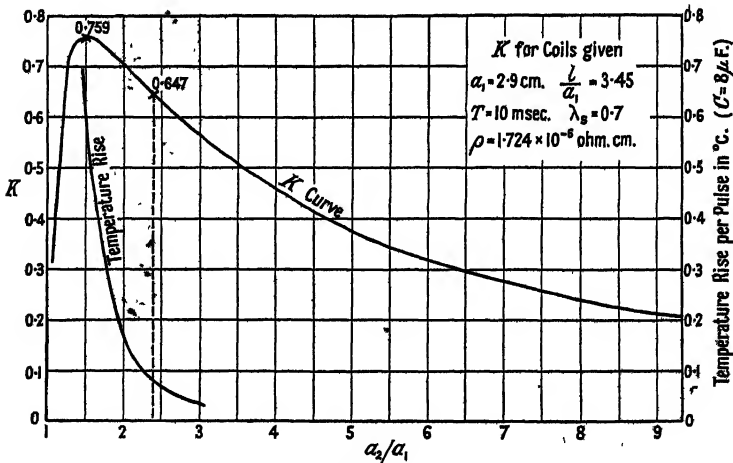


Figure 7.

Note. In this Figure $T=10$ msec. should read $T=10.7$ msec.

20 cm. and 17 s.w.g. is very slightly thicker than required. With this coil $H=21,200$ gauss with $C=7 \mu F.$ and $V=25,000$ volts.*

* Since the coil was constructed the condenser bank has been increased to $8 \mu F.$, giving $H=22,700$ gauss and $T=10.7$ msec.

Mechanical Stresses in Coil

With intense magnetic fields the stresses in a coil may become very large. Cockcroft (1928) has given formulae and data which enable these stresses to be calculated. The main resultant force is perpendicular to the axis and outwards (compensated by binding the coil) but with very intense fields (say, of the order of 500 kilogauss) the forces on the inner parts of the coil may be so great as to cause the copper to flow towards the outer surface.

The maximum values of the stresses occur at the mid-point of the external surface and, with the coil designed above, are 6.9 kg/cm² in a direction perpendicular to the axis and 4.3 kg/cm² parallel to it. The total bursting force would be approximately 4,800 kg. The copper windings should stand this force, but, nevertheless, the coil is reinforced with eight $\frac{1}{2}$ -inch thick radial Tufnol strips, securely held by the end cheeks, pressed against the outer surface of the coil and with a brass band round their outer edges. The bolts securing the brass band are insulated from one end to minimize eddy currents. The Tufnol strips are set radially to facilitate efficient cooling of the surface of the coil.

§ 4. K AS A FUNCTION OF ITS PARAMETERS

K may be regarded as an efficiency constant for the operation of a coil in the way specified at the beginning of this article. However, it must be pointed out that the field produced also depends on the energy stored and the volume enclosed by the inner diameter of the coil.

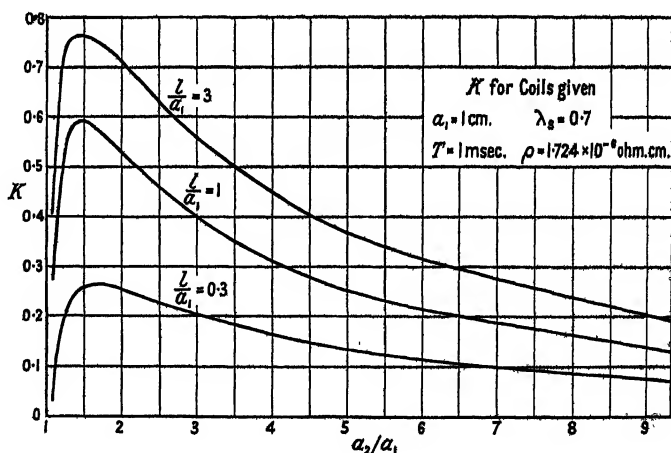


Figure 8.

Figure 8 gives examples of the way in which K (as a function of a_2/a_1) depends on l/a_1 . The value of a_2/a_1 for maximum K does not change much but, as would be expected, K is much smaller for shorter coils. In Figure 9 can be seen the very great change in K produced by changes in a_1 . This arises from the fact that, with smaller a_1 , the resistance of the coil becomes much more appreciable and, not only is K much smaller, but the value of a_2/a_1 at which the maximum of K occurs is larger. Figure 10 illustrates the way in which K depends on T .

Now, the only way in which K depends on parameters other than those of shape (i.e. l/a_1 and a_2/a_1) is in the expression for RT/L (of which J is a function) :

$$\frac{RT}{L} = \frac{\rho 10^9 T}{a_1^2 \lambda_s \phi}.$$

The effect due to a_1 varies inversely as its square, whereas that due to T is proportional to its first power. From the expression for RT/L we see that

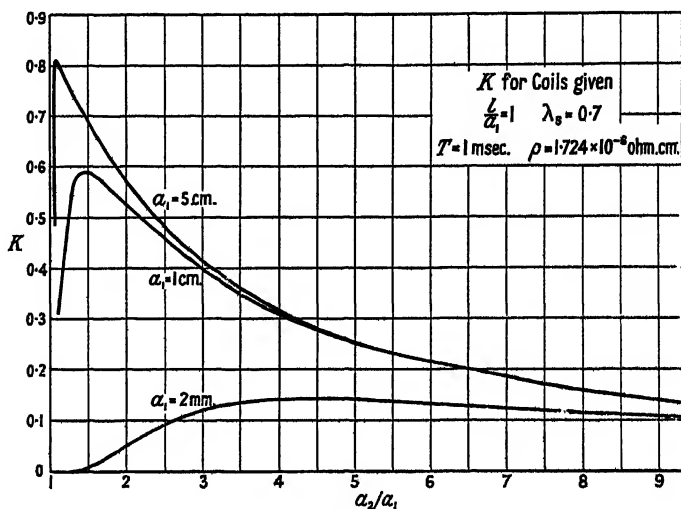


Figure 9.

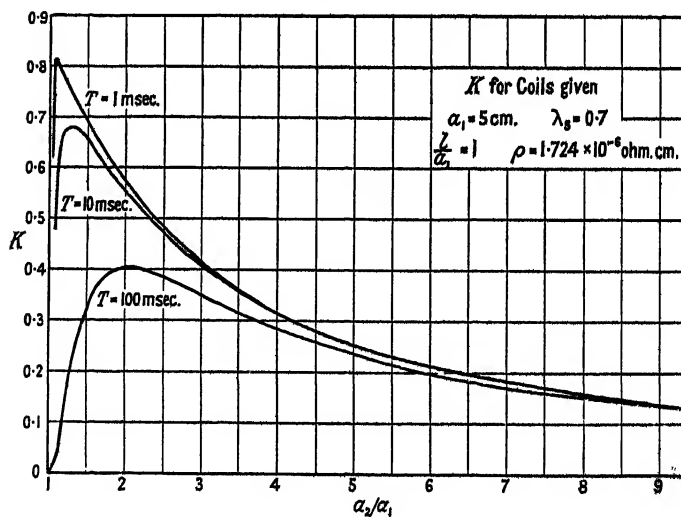


Figure 10.

any proportional change in ρ or $1/\lambda_s$ has the same effect on the value of K as the same proportional change in T .

The curves in Figures 8, 9, 10 and the above considerations should be of use when a coil is to be designed, making it possible to assess rapidly the influence of the various parameters on the magnetic field produced.

§ 5. SPATIAL VARIATIONS OF THE MAGNETIC FIELD

In many investigations one other property is important, viz. the degree of uniformity of the magnetic field. As the requirements of uniformity depend very much on the particular investigation, no general criterion will be presented, but merely data which make it possible to calculate the degree of uniformity of the field produced by a particular coil. As would be expected, the relative values of the field at various points near a coil depend only on the shape of the coil.

The value of the field at the centre of a coil is given by equation (4) and the value at a point on the axis distant x from the centre by

$$H_{0x} = \frac{2\pi nI}{20l(a_2 - a_1)} \times \left[(l+x) \ln \frac{a_2 + \{(l+x)^2 + a_2^2\}^{1/2}}{a_1 + \{(l+x)^2 + a_1^2\}^{1/2}} + (l-x) \ln \frac{a_2 + \{(l-x)^2 + a_2^2\}^{1/2}}{a_1 + \{(l-x)^2 + a_1^2\}^{1/2}} \right]. \quad \dots\dots(7)$$

If (7) is divided by (4) and then both numerator and denominator of the right-hand side of the resulting expression are divided by a_1 , H_{0x}/H_{00} is obtained as a function of a_2/a_1 , l/a_1 and x/a_1 . Thus, if in a design problem, instead of l/a_1 being specified, H_{0x}/H_{00} is specified for a given value of x/a_1 then, from equations (4) and (7), the value of l/a_1 for each a_2/a_1 can be calculated. However, the value of H_{0x}/H_{00} is almost independent of a_2/a_1 and thus it will often be possible to obtain the appropriate value of l/a_1 for one value of a_2/a_1 , and then use it with all required values of a_2/a_1 .

In general the magnetic field at points off the axis is not in the direction of the axis, but, with the rotationally symmetrical coils being considered, it will only have components parallel to the axis and in a radial direction. The general expressions for these components in terms of the value of the field at the axis are (Cosslett 1946) :

$$\text{In the direction of } x : H_{rx} = H_{0x} - \frac{r^2}{4} H_{0x}'' + \frac{r^4}{64} H_{0x}^{(4)} \dots \frac{(-1)^n}{(n!)^2} \left(\frac{r}{2}\right)^{2n} H_{0x}^{(2n)} \quad \dots\dots(8)$$

$$\text{and in the direction of } r : H_{rx} = -\frac{r}{2} H_{0x}' + \frac{r^3}{16} H_{0x}''' \dots \frac{(-1)^n}{n!(n-1)!} \left(\frac{r}{2}\right)^{2n-1} H_{0x}^{(2n-1)} \quad \dots\dots(9)$$

The differentiation is with regard to x . The coefficients can be obtained from equation (7) by differentiation but, in general, the full expressions for these and, consequently, the components of the field involve a very large number of terms. However, the case of the central plane of the coil ($x=0$) is much simpler. As the coils are symmetrical with regard to their central plane the value of the radial component of H_{r0} is zero for all values of r . For small values of r the axial component may be represented by

$$H_{r0} = H_{00} - \frac{r^2}{4} H_{00}''$$

where H_{00} is given by equation (4), and, by repeated differentiation of equation (7) and putting $x=0$,

$$H_{00}'' = \frac{2\pi nI}{10(a_2 - a_1)} \left[\frac{3}{(a_2^2 + l^2)^{1/2} \{a_2 + (a_2^2 + l^2)^{1/2}\}} - \frac{3}{(a_1^2 + l^2)^{1/2} \{a_1 + (a_1^2 + l^2)^{1/2}\}} \right. \\ \left. - \frac{l^2}{(a_2^2 + l^2)^{3/2} \{a_2 + (a_2^2 + l^2)^{1/2}\}} + \frac{l^2}{(a_1^2 + l^2)^{3/2} \{a_1 + (a_1^2 + l^2)^{1/2}\}} \right. \\ \left. - \frac{l^2}{(a_2^2 + l^2) \{a_2 + (a_2^2 + l^2)^{1/2}\}^2} + \frac{l^2}{(a_1^2 + l^2) \{a_1 + (a_1^2 + l^2)^{1/2}\}^2} \right]. \quad \dots\dots(10)$$

ACKNOWLEDGMENTS

The author wishes to acknowledge help and suggestions given in this work by Professor J. Sayers and assistance given by Mr. N. L. Allen. He is indebted to the English Electric Co. for the gift of an experimental glass ignitron which formed a vital part of the electronic switch. This work was made possible through the award to the author of a British Council Scholarship.

Note added in proof. Since the completion of this paper, a paper by G. Raoult (1949) has come to the author's attention. It is entitled "Réalisation de champs magnétiques intense par impulsions. Applications aux phénomènes de polarisation rotatoire et de biréfringence magnétique". In this, besides other work, Raoult describes coils, with the requisite theory, for producing half-wave intense magnetic field pulses. His theory for the design of coils is somewhat similar to that above but, in his final expressions, he has neglected the effect of the resistance of the coil on the peak current obtained. In other words, he has assumed that $J=1$. In Raoult's work $k_1 = (20a_1/l)^{1/2} 10^8 S$ and $D = 1/10\lambda^{1/2}$.

REFERENCES

- COCKCROFT, J. D., 1928, *Phil. Trans. Roy. Soc. A*, **227**, 317.
 COSSLETT, V. E., 1946, *Introduction to Electron Optics* (Oxford: Clarendon Press), p. 106.
 GROVER, F. W., 1922, *Scientific Papers of the Bureau of Standards*, **18**, 415.
 RAOULT, G., 1949, *Ann. Phys., Paris*, 12^e serie, **4**, 369.

A Theory of Contact Noise in Semiconductors

By G. G. MACFARLANE

Telecommunications Research Establishment, Ministry of Supply

MS. received 8th May 1950

ABSTRACT. A theory of contact noise is described in which the low-frequency noise is attributed to the random movement of adsorbed ions on the surface of a semiconductor from which an electron current is being drawn. Emission of electrons is assumed to take place only at localized patches on the surface and the adsorbed ions are assumed to give rise to a Schottky barrier layer, in which the potential maximum is linearly related to the concentration of ions. Diffusion of the ions over the surface gives rise to random fluctuations in the concentration of ions in a patch which results in random fluctuations in the height of the potential barrier and the emission current. It is shown that for a circular patch the spectral power density of the noise current varies with mean current j_0 and frequency f as $j_0^x f^x$ over a small range of frequency and that x varies monotonically from -0.75 at the lowest frequencies to -1.125 at the highest frequencies. It is also shown that for a long thin rectangular patch the index x varies monotonically from -0.5 to -1.5 as the frequency is increased from zero. The dependence of the noise power density on temperature is also discussed.

§ 1. INTRODUCTION

IN a previous paper (Macfarlane 1947) the author has described a theory of contact noise which depends on a diffusion conduction process in the atomic layers next to the emitting surface of a crystal. The noise was attributed to the diffusion of clusters of mobile atoms on to the contact surface, their ionization on the surface and subsequent conduction away from the surface in the applied field. On the basis of a simple model it was shown that the spectral power density of the current noise should depend on mean current j_0 and frequency f as $j_0^{x+1} f^{-x}$. The index x , which lies between 1 and 2, depends on the concentration N of atoms in a mobile cluster. As x tends to 1, N tends to ∞ . However, experiments on silicon crystal rectifiers (Miller 1947), carbon granule microphones (Christensen and Pearson 1936, Miller 1949), germanium crystal rectifiers and lead sulphide photosensitive layers (Harris, Abson and Roberts 1947) show that in some cases x is very nearly equal to unity and in other cases it must be taken less than unity. In addition some recently published measurements (Campbell and Chipman 1949) on resistors have shown that the index of f varies from about -1 at low frequencies to -1.6 at high frequencies. In order to explain these results on the above theory it would be necessary to assume enormous concentrations of atoms in mobile clusters. It is therefore apparent that this theory, in spite of its success in giving the observed dependence on current and frequency with micro-crystalline layers of PbS (Harris, Abson and Roberts 1947), is inadequate.

In order to overcome these difficulties the author has developed a new theory, in which the noise is attributed to the random modulation of the Schottky barrier potential at emitting patches by random fluctuations in the concentration of mobile adsorbed ions. In its simplest form the theory refers to thermionic emission from a semiconductor, but it can be applied equally well to emission from one crystallite to another at contacts.

The idea is that the surface of a semiconducting crystal is covered by a partial film of adsorbed atoms, which have become negative ions by the capture of electrons from the crystal. As a result a Schottky space charge layer is set up at the surface, as in Bardeen's (1947) theory of surface states. The height of the potential barrier is proportional to the concentration of ions. The adsorbed ions are assumed to move at random over the surface of the crystal. When an electric field is applied to the crystal, emission of electrons is assumed to take place only at localized patches on the surface. Fluctuations in the concentration of ions in a patch occur due to the diffusion of ions out of and into the patch. These fluctuations in number of ions in a patch give rise to fluctuations in the height of the potential barrier at the patch and therefore to fluctuations in emission of electrons.

Since the electron emission is controlled by the concentration of ions in a patch the problem is to study the correlation in the number of ions in a patch and its dependence on time.

The problem is essentially the same as a classical one in the study of the Brownian movement in colloidal solutions. It is to find the correlation function for the fluctuation in the number of particles in a cylindrical volume within a much larger volume of colloidal solution. An account of the relevant theory of colloidal statistics is given by Chandrasekhar (1943).

In applying the formulae for contact noise to a multicrystalline layer or mass of semiconducting crystals it must be clearly realized that there are also other sources of noise, such as the Johnson noise of the bulk resistance, fluctuations in the number of current carriers, and noise due to the 'spreading' resistance. The power density of these other sources is, however, independent of frequency at least at low frequencies, whereas contact noise per unit bandwidth increases as the frequency is decreased. If a frequency power spectrum of the total noise is known it is only necessary to subtract out the constant noise power, observed at high frequencies, in order to obtain the spectrum of the contact noise.

In a very recent article, which came to the notice of the author after the theory to be described below had been worked out, Richardson (1950) has described in a most interesting manner a theory of contact noise in which the noise is also attributed to a diffusing adsorbed layer. He considers the region of multiple contact between two rough surfaces and assumes that the conductance per unit area is a function of the separation of the surfaces and the total concentration of adsorbate, near the point in question. By an analysis quite different from the one given below he derives an expression for the spectral power density which varies with frequency as f^{-1} . Since the total noise power would diverge to infinity with this law Richardson suggests a modification to remove the divergence. With our model we arrive at a different expression which does not suffer from this defect. As we shall show it gives a law of the form $j_0^2 f^\alpha$ over a small frequency range, where the index α varies from -0.75 at very low frequencies to -1.125 at very high frequencies for a circular patch and from -0.5 to -1.5 for a long thin rectangular patch.

§2. ANALYSIS OF THE PATCH MODEL

We consider emission of electrons from a localized surface patch A of an emitter. Emission from the surface immediately surrounding A is assumed to be negligible. We assume that the current j through the patch A depends on the

number n of adsorbed ions within the patch as $\exp(an)$ and that the adsorbed ions can move at random over the entire surface of A and the surrounding surface.

With these assumptions we can calculate the auto-correlation function and from it, by Khintchine's theorem, the spectral power density of the noise current emitted by the patch in the following manner.

The first step is to calculate the mean number of ions \bar{m} in a patch A after time τ from the instant when n ions occur in A . In doing this we assume that the motions of the ions are independent of each other and that all positions in A have equal *a priori* probability. Then if the probability that an ion somewhere in A will have emerged from it during the time τ is P , the number of ions which on the average will leave the patch in τ is nP and therefore the number remaining is $n(1-P)$. In addition the number entering A on the average in τ is NP regardless of the initial value of n . The average number of ions in the patch after time τ is therefore

$$\bar{m} = n(1-P) + NP. \quad \dots\dots(1)$$

Now in accordance with our assumptions the current j through the patch A depends on n as $j = j_0 e^{a(n-N)}$, where j_0 is the mean current. The noise current is therefore

$$\Delta j = j - j_0 = j_0 [e^{a(n-N)} - 1] \simeq j_0 a(n-N). \quad \dots\dots(2)$$

The approximation used in (2) is justified provided the noise current is small compared with the mean current.

The auto-correlation function of the noise current is therefore

$$\begin{aligned} f(\tau) &= \overline{\Delta j(n) \cdot \Delta j(m)} \\ &= \overline{(j_0 a)^2 (n-N)(m-N)} \\ &= \overline{(j_0 a)^2 (n-N)^2} (1-P) \\ &= (j_0 a)^2 N(1-P). \end{aligned} \quad \dots\dots(3)$$

In the third line of (3) we have averaged with respect to m keeping n constant and using (1), and in the fourth line we have used the well-known result that the mean square deviation for random events is equal to the mean.

The next step is to consider the probability P . Suppose firstly that the patch A is a circular disc of radius r and that the diffusion constant for the movement of ions on the surface is D , then it can be shown (Chandrasekhar 1943) that the probability P that an ion somewhere within A will have emerged from it during the time τ is

$$P = e^{-x} [I_0(x) + I_1(x)], \quad \dots\dots(4)$$

where

$$x = (\tau_0/\tau)^{1/2} \quad \dots\dots(5)$$

and

$$\tau_0 = (2r)^2/D. \quad \dots\dots(6)$$

Using Khintchine's theorem (Wang and Uhlenbeck 1945) we can now write down an expression for the spectral power density of the noise current.

$$R(\omega) = 4 \int_0^\infty f(\tau) \cos(\omega\tau) d\tau = 4(j_0 a)^2 N \tau_0 F(p) \quad \dots\dots(7)$$

$$\text{where } F(p) = \int_0^\infty \{1 - \exp(-x^{-1/2}) [I_0(x^{-1/2}) + I_1(x^{-1/2})]\} \cos(px) dx \quad \dots\dots(8)$$

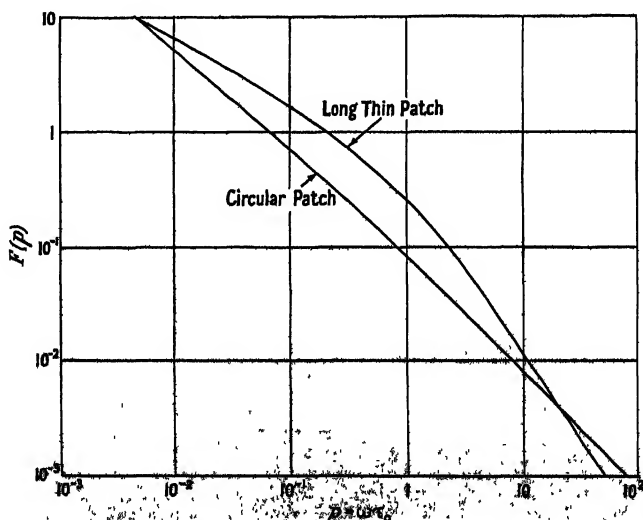
$$\text{and } p = \omega\tau_0. \quad \dots\dots(9)$$

The function $F(p)$ has been evaluated from the series given in the Appendix and is shown in the Figure. Two limiting formulae are useful.

When $p \ll 1$ $F(p) \simeq 0.2345p^{-3/4}$ (10)

and when $p \gg 1$ $F(p) \simeq 0.1466p^{-9/8}$(11)

In the preceding analysis $R(\omega)$ is the mean square noise current at a single localized patch or contact. It was derived on the assumption of a constant voltage across the contact. When there are many such contacts, as in a multi-crystalline resistor or at contacts between rough surfaces, the noise fluctuation of the total input current is a weighted sum of the separate fluctuations at the contacts. The weighting factor will depend on the current at the contact and on the current



which a series R.M.F. at the contact would produce at the input to the resistor. It will therefore depend on the position of the contact in the resistor.

As an example consider a resistor which can be represented by a set of contacts in parallel, each set consisting of a contacts in series. Between adjacent series contacts assume a 'bulk' resistance. Let the resistance of the p th contact be r_{pq} and the 'bulk' resistance in series be R_{pq} . Then a simple circuit analysis shows that the mean square noise current at input is

$$\overline{\Delta j^2} = j_0^2 R^2 \sum_{p=1}^m \sum_{q=1}^n c_{pq} r_{pq}^2 / R_p^4, \quad \text{.....(12)}$$

where j_0 is the mean current at input, R the total resistance of the resistor, $R_p = \sum_{q=1}^n (R_{pq} + r_{pq})$ and

$$c_{pq} = 4a^2 N_{pq} \tau_{pq} F(\omega \tau_{pq}). \quad \text{.....(13)}$$

As a further simplification suppose that all the contacts have the same resistance r_0 and all the resistances R_{pq} are equal to R_0 . Then

$$\overline{\Delta j^2} = j_0^2 \{r_0 / (R_0 + r_0)\}^2 \bar{c} / mn. \quad \text{.....(14)}$$

Although this last expression applies only to a very simple resistor it does indicate that the power density of the noise current decreases as the number of contacts is increased.

On account of the summation in (12) the variation of noise power density with frequency will not be quite given by (8). However, at high enough frequencies where (11) is valid, $c_{pq} \sim 0.5864 a^2 N_{pq} \tau_{pq}^{-1/8} \omega^{-9/8}$ and $\overline{\Delta j^2}$ is then directly proportional to $\omega^{-9/8}$. Similarly at low enough frequencies $\overline{\Delta j^2}$ varies with frequency as $\omega^{-3/4}$.

§3. THE EFFECT OF PATCH SHAPE ON THE SPECTRUM

In §2 we assumed a circular patch. However the shape of a patch does have an interesting effect on the spectrum, as we can see by evaluating $R(\omega)$ for a long thin strip. Then diffusion takes place primarily across the strip. If the width of the strip is h the probability P is given by (Chandrasekhar 1943)

$$1 - P = \text{erf}(x^{-1/2}) - (x/\pi)^{1/2} [1 - e^{-1/x}], \quad \dots\dots(15)$$

where $x = \tau_0/\tau$ and $\tau_0 = h^2/4D$.

The spectral power density of the noise current is then

$$R(\omega) = 4(j_0 a)^2 N \tau_0 F(p), \quad \dots\dots(16)$$

where $p = \omega \tau_0$ and

$$F(p) = \int_0^\infty \{ \text{erf}(x^{-1/2}) - (x/\pi)^{1/2} [1 - e^{-1/x}] \} \cos(px) dx. \quad \dots\dots(17)$$

The function $F(p)$ has been evaluated from the formulae

$$F(p) = \sum_{n=1}^{\infty} \frac{(-)^n \sqrt{\pi} p^{n-3/2}}{4\Gamma(n+1)\Gamma(n+\frac{1}{2}) \sin \pi(\frac{1}{2}n - \frac{1}{4})} + \sum_{n=0}^{\infty} \frac{(-)^{n+1} \sqrt{\pi} p^{2n}}{2\Gamma(2n+2)\Gamma(2n+5/2)} \quad \dots\dots(18)$$

$$\simeq (2p)^{-1/2} \text{ for } p \ll 1 \quad \dots\dots(19)$$

$$\text{and} \quad \simeq (2p)^{-3/2} \text{ for } p \gg 1. \quad \dots\dots(20)$$

These formulae were derived by the method given for (8) in the Appendix. $F(p)$ for a long thin strip is shown plotted in the Figure for comparison with $F(p)$ for a circular patch.

§4. DISCUSSION OF THE NOISE SPECTRUM

Consider firstly the spectrum for a circular patch. If $F(p)$ is replaced by a power law cp^x over a small range of p then it can be seen from formulae (10) and (11) and from the Figure that x varies from -0.75 when $p \ll 1$ to -1.125 when $p \gg 1$. From this it follows that the integral of $F(p)$ taken over the entire frequency range from $p=0$ to $p=\infty$ is finite, notwithstanding the fact that $F(p)$ tends to infinity as p tends to zero.

It will be observed that the spectral power density of the noise current increases with the square of the mean current and directly with the average number of adsorbed ions in a patch.

Only an approximate expression for the dependence of $R(\omega)$ on temperature can be given unless the variation with temperature of the mean number, N , of ions in a patch is known. If we assume N independent of temperature T we have

$$a = E_0/kT, \quad D = D_0 e^{-E/kT}, \quad \tau_0 = (2r)^2 D_0^{-1} e^{E/kT}, \quad \dots (21)$$

where E is the activation energy for diffusion of ions over the surface and E_0 is the change in height of the barrier potential at a patch produced by one ion. Then, if $\omega\tau_0 \gg 1$, we get from (7) and (11) that

$$R(\omega) \sim 0.1868 \pi j_0^2 N D_0^{1/2} (2r)^{-1/4} (E_0/kT)^2 e^{-E/8kT} \omega^{-9/8} \\ \propto T^{-2} e^{-E/8kT}. \quad \dots (22)$$

From (22) it follows that $R(\omega)$ increases with T for temperatures below $T_c = E/16k$ and decreases with increase of T for temperatures above T_c . If $E \approx 0.1$ ev., which is probably rather a high value for surface diffusion, $T_c \approx 60^\circ \text{K}$. Then for temperatures above 60°K . and for high frequencies $R(\omega)$ should decrease as the temperature is increased. Evidence that $R(\omega)$ does decrease with increase of temperature is provided by measurements on silicon rectifiers (Miller 1949).

Although little is known of the values of the diffusion constants D_0 and E , for diffusion of adsorbed ions over a crystal face both D_0 and E are likely to be much smaller than for diffusion of impurities through a crystal (Mott and Gurney 1948). In order to appreciate the magnitude of τ_0 take $D_0 = 10^{-4} \text{ cm}^2/\text{sec.}$, $E = 0.1$ ev., and take the radius of a patch as 10^{-4} cm . Then $\tau_0 \approx 0.016 \text{ sec.}$ at room temperature.

Consider now the effect of patch shape on the spectrum. We have found that, when $\omega\tau_0 \gg 1$, $R(\omega) \propto \omega^{-9/8}$ for a circular patch and $R(\omega) \propto \omega^{-5/2}$ for a long thin rectangular patch and that, when $\omega\tau_0 \ll 1$, $R(\omega) \propto \omega^{-3/4}$ for a circular patch and $R(\omega) \propto \omega^{-1/2}$ for a long thin rectangular patch. Deviations in the index x in the expression $R(\omega) \propto \omega^x$ can therefore be explained in terms of patch shape so long as x lies in the range $-1.125 > x > -1.5$ at high frequencies and in the range $-0.75 < x < -0.5$ at the lowest frequencies. It has been found experimentally (Miller 1949) that for a photoelectric cell illuminated with a constant intensity of light $x \approx -1$, for a carbon granule microphone $-1 > x > -1.1$, and for a carbon resistor $-1.1 > x > -1.2$. These results can therefore be explained by assuming nearly circular patches. For various types of resistor Campbell and Chipman (1949) have found that the index x varies from about -1 at a frequency of 20 kc/s. to about -1.6 at 500 kc/s. and that $R(\omega)$ increases as the square of the mean current. It would therefore be necessary to assume long thin patches in order to explain these results. For micro-crystalline layers of PbS it is found (Harris, Abson and Roberts 1947) that x may lie between -1 and -1.4 .

It appears therefore that the theory of contact noise described above, in which the noise is attributed to fluctuations in emission from localized patches on the surface of an emitter due to fluctuations in the number of ions in a patch, can provide a reasonable explanation of the observed results.

APPENDIX

EVALUATION OF THE INTEGRAL FOR $F(p)$

The Mellin transform of $1 - P(x) = 1 - \exp(-x^{-1/4})[I_0(x^{-1/4}) + I_1(x^{-1/4})]$ is (Macfarlane 1949)

$$F_1(s) = \frac{-2^{4s+2}\Gamma(\frac{1}{2}+4s)\Gamma(-4s)}{\sqrt{\pi}\Gamma(2+4s)} \quad 0 < \sigma < \frac{1}{4}.$$

By using the inversion formula of the Mellin transform to express $1 - P$ as an integral in the s -plane and reversing the order of integration we get

$$\begin{aligned} \int_0^\infty [1 - P(x)] \cos(px) dx &= \frac{1}{2\pi i} \int_{\sigma-i\infty}^{\sigma+i\infty} F_1(s) ds \int_0^\infty x^{-s} \cos(px) dx \quad 0 < \sigma < \frac{1}{4} \\ &= \frac{-1}{2\pi i} \int_{\sigma-i\infty}^{\sigma+i\infty} \frac{4\Gamma(-4s)\Gamma(\frac{1}{2}+4s)\Gamma(\frac{1}{2}-\frac{1}{2}s)2^{3s}}{\Gamma(2+4s)\Gamma(\frac{1}{2}s)} p^{s-1} ds \quad 0 < \sigma < \frac{1}{4}. \end{aligned}$$

The integrand has poles at

$$s = -\frac{1}{4}(n + \frac{1}{2}) = -\frac{1}{8}, -\frac{3}{8}, -\frac{5}{8}, \dots \text{ due to } \Gamma(\frac{1}{2} + 4s)$$

$$\text{and at } s = n/4 \ (n > 0) = \frac{1}{4}, \frac{1}{2}, \frac{3}{4}, \dots \text{ due to } \Gamma(-4s)$$

and double poles at

$$s = 2n + 1 = 1, 3, 5, \dots \text{ due to } \Gamma(-4s) \text{ and } \Gamma(\frac{1}{2} - \frac{1}{2}s).$$

Moving the contour steadily to the right we can take out the residues at the poles to the right of σ and get the following convergent series for $F(p)$.

$$\begin{aligned} F(p) &= \sum_{\substack{n=1 \\ n \neq 8r+4}}^\infty \frac{\sqrt{\pi}(-)^{n+1}\Gamma(n+\frac{1}{2})2^{n-1}p^{n/4-1}}{\Gamma(n+1)\Gamma(n+2)\Gamma(\frac{1}{4}n)\cos(n\pi/8)} \\ &+ \sum_{n=0}^\infty \frac{(-)^n\Gamma(8n+\frac{3}{2})2^{8n+4}p^{2n}}{\sqrt{\pi}\Gamma(8n+6)\Gamma(8n+5)\Gamma(2n+1)} \left[4\psi(8n+\frac{3}{2}) - 4\psi(8n+6) - 4\psi(8n+5) \right] \\ &\quad - \psi(2n+1) + \ln(16p) \end{aligned}$$

The first few terms are

$$\begin{aligned} &0.23447p^{-3/4} - 0.31333p^{-1/2} + 0.34892p^{-1/4} - 0.14378 + 0.03646 \ln p \\ &- 0.04953p^{1/4} + 0.00718p^{1/2} - 0.00123p^{3/4} + \dots \end{aligned}$$

Moving the contour to the left we obtain an asymptotic series for $F(p)$ from the residues at the poles of $\Gamma(\frac{1}{2} + 4s)$.

$$\begin{aligned} F(p) &\sim \sum_{n=0}^\infty \frac{(-)^n\Gamma(n+\frac{1}{2})\Gamma(n-\frac{1}{2})\Gamma(\frac{1}{4}n+\frac{3}{8})}{\pi^{3/2}\Gamma(n+1)} \sin \pi (\frac{1}{8}n + \frac{1}{16}) 2^{-n-1/2} p^{n/4-3/8} \\ &\sim 0.14659p^{-9/8} - 0.04925p^{-11/8} - 0.01394p^{-13/8} - 0.01093p^{-15/8} \\ &- 0.01328p^{-17/8} \dots \end{aligned}$$

ACKNOWLEDGMENT

The author desires to thank the Ministry of Supply for permission to publish his paper, which is reproduced with the permission of the Controller, I.M. Stationery Office.

REFERENCES

- BARDEEN, J., 1947, *Phys. Rev.*, **71**, 717.
 CAMPBELL, R. H., and CHIPMAN, R. A., 1949, *Proc. Inst. Radio Engrs.*, **37**, 938.
 CHANDRASEKHAR, S., 1943, *Rev. Mod. Phys.*, **15**, 1, Chap. III.
 CHRISTENSEN, C. J., and PEARSON, G. L., 1936, *Bell Syst. Tech. J.*, **15**, 197.
 HARRIS, E. J., ABSON, W., and ROBERTS, W. L., 1947, T.R.E. Report T2051 (unclassified).
 MACFARLANE, G. G., 1947, *Proc. Phys. Soc.*, **59**, 366; 1949, *Phil. Mag.*, Ser. 7, 188.
 MILLER, P. H., 1947, *Proc. Inst. Radio Engrs.*, **35**, 252; 1949, See review article prepared by S. J. ANGELLO, *Electrical Engineering*, **68**, 870.
 MOTT, N. F., and GURNEY, R. W., 1948, *Electronic Processes in Ionic Crystals*, 2nd Edition. (Oxford: Clarendon Press).
 RICHARDSON, J. M., 1950, *Bell Syst. Tech. J.*, **29**, 117.
 WANG, M. C., and UHLENBECK, 1945, *Rev. Mod. Phys.*, **17**, 323.

Dielectric Loss and Dielectric Constant Measurements in Supercooled Liquids

BY C. DODD AND G. N. ROBERTS
 University College and Queen Mary College, London

Communicated by E. N. da C. Andrade; MS. received 26th June 1950

ABSTRACT. In order to provide further evidence in support of the structural change which takes place in a liquid when it is supercooled, measurements of dielectric loss at high frequencies and of dielectric constant have been made on various polar liquids. The evidence from the dielectric loss measurements is inconclusive but for the four liquids tested a significant discontinuity in the dielectric constant-temperature curve has been established at the melting point in each case.

§ 1. INTRODUCTION

THE work of Dodd and Hu Pak Mi (1949) has shown for certain polar liquids that, although there is no discontinuity of the viscosity η at the melting point, there is a sharp change in the course of the line of $\log \eta$ plotted against $1/T$ at this point. Such a phenomenon can only result from some sort of structural change in the liquid as it supercools and any such change should also show itself in the temperature variation of other properties of the liquid.

The present paper describes measurements made on the dielectric loss and dielectric constant of several liquids over a range of temperature on both sides of the melting point.

§ 2. DIELECTRIC LOSS

The apparatus for the determination of the dielectric loss was one in use in the research department of Electrical Engineering at Queen Mary College, London. Oscillations of wavelength 3 cm. from a klystron were fed into an H_{01} cavity resonator containing the liquid, and were passed through coaxial cable to a crystal detector, variable attenuator and galvanometer. The resonator was in a constant temperature bath and a thermocouple in contact with the movable piston attached

to a micrometer shaft registered the temperature of the liquid. The width of the resonance curve was obtained from the positions of the micrometer shaft when the galvanometer deflections were half the maximum value. From this width the dielectric loss was calculated. In Figure 1 the variation of loss with temperature is shown by plotting ϵ''/ϵ_0 against temperature for both normal and supercooled liquid phenyl ether (M.Pt. 26.9°C .) where ϵ'' is the imaginary part of the complex dielectric constant and ϵ_0 the dielectric constant of free space. It is seen that over this range the scatter of the points is too great for any indication of a discontinuity at the melting point to be apparent.

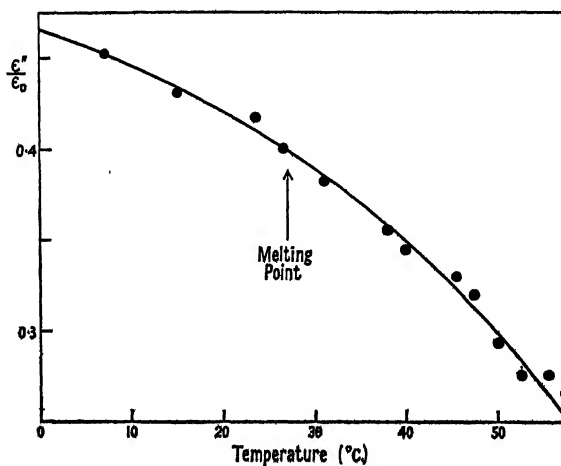


Figure 1.

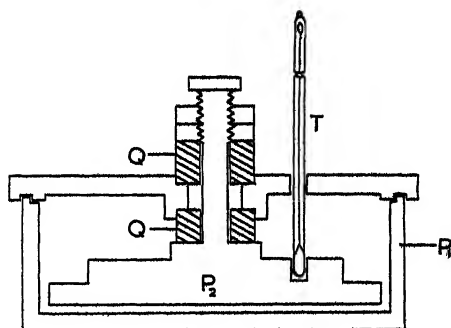


Figure 2.

§ 5. DIELECTRIC CONSTANT

In order to measure small changes in the dielectric constant of a liquid to a high degree of accuracy, a test condenser was joined in parallel with a standard and a variable condenser in a circuit tuned to a frequency of 244 kc/s.

The test condenser (Figure 2) was of the parallel plate type, the plates P_1 , P_2 of diameter about 6 cm. being separated by a distance of the order of 1 mm. by the quartz washers QQ. The empty capacity was determined at various temperatures by a modified Hartshorn circuit. This value was about 28 pF. and changed by only 0.18 pF. for a temperature change of 40°C . The capacity varied regularly with temperature and showed no hysteresis when taken over a heating cycle. The

whole cell was then filled with liquid, the thermometer T registering the temperature to an accuracy within 0.02°C. , and was immersed in a bath whose temperature could be maintained constant to better than 0.01°C.

The variable condenser was of the cylindrical type, the capacity being varied by movement of the inner cylinder. The total movement available was 1,500 divisions on the micrometer, one division (10^{-3} cm.) corresponding to a capacity change of 0.01074 pF. A change of half a division caused a measurable change in the resonant position.

The setting of the micrometer condenser to give resonance was found with the test condenser at various temperatures both above and below the melting point of the liquid. The liquids chosen, which were all polar, readily supercooled and remained in the liquid state throughout. The micrometer condenser setting C is plotted against temperature for phenyl ether, M.Pt. 26.9°C. , $\mu = 1.17$ (Figure 3); salol, M.Pt. 42°C. , $\mu = 3.15$ (Figure 4); menthol, M.Pt. 42°C. , $\mu = 1.56$ (Figure 5); and azoxybenzene, M.Pt. 36°C. , $\mu = 1.70$ (Figure 6); μ is the dipole moment, in all cases $\times 10^{18}$. Because of the large ionic loss for azoxybenzene at the lower frequency (244 kc/s.), the circuit was tuned to a frequency of 1.65 Mc/s. for this liquid. Values for the dielectric constant K deduced from the corresponding micrometer condenser settings, are given as ordinates on the right-hand side of the graphs.

§ 4. DISCUSSION

Examination of Figures 3, 4, 5 and 6 shows in each case that for the liquid above its melting point the points lie fairly well on a straight line. Those for the super-

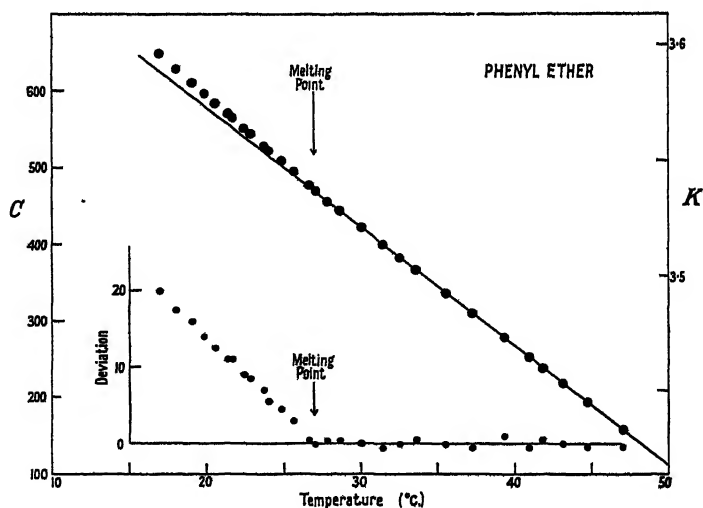


Figure 3.

cooled liquid also lie on a line which is, however, inclined to the line for the normal liquid. In each case these lines intersect in the neighbourhood of the melting point. This discontinuity at the melting point is brought out more strikingly by the insets to the graphs, where the deviations of *all* points from the straight line, fitted by least squares, for the liquid above the melting point, are plotted against

temperature. It should be mentioned that in order to avoid any systematic error arising in the apparatus, the dielectric constant was measured first in the region above the melting point, then in the slightly supercooled region, back into the normal region, further into the supercooled region, and so on.

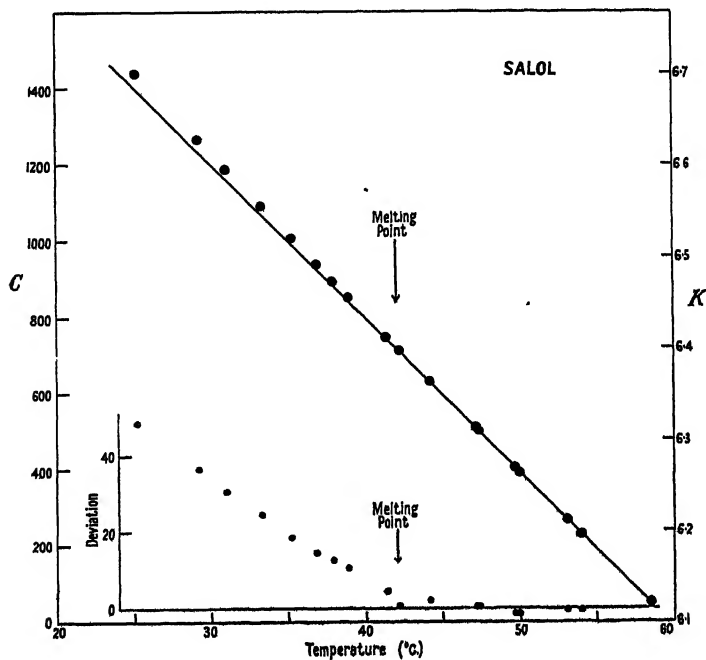


Figure 4.

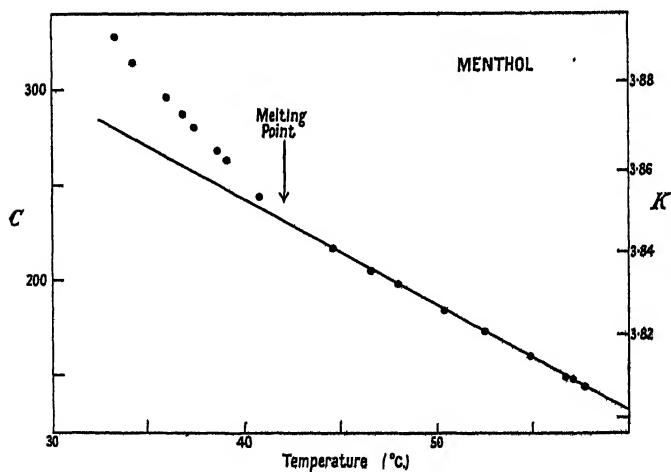


Figure 5.

For phenyl ether, salol and menthol similar discontinuities at their melting points have been found by Dodd and Hu (1949) for the variation of viscosity with temperature. The experimental technique used for these viscosity measurements

is not, however, easily adaptable to azoxybenzene, which is a dark, almost opaque liquid.

It would therefore appear that the change occurring in the nature of the liquid as it enters the supercooled region is responsible for both the discontinuity in

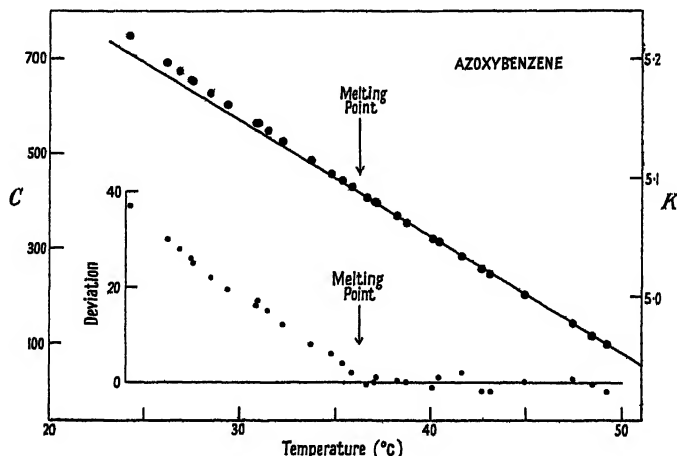


Figure 6.

dielectric constant and the discontinuity in viscosity. The search for similar discontinuities in other physical properties of these supercooled liquids is continuing.

ACKNOWLEDGMENTS

We should like to express our gratitude to Professor W. J. John of the Electrical Engineering Department, Queen Mary College, for permission to use apparatus in his department, and to the Department of Scientific and Industrial Research for a grant to one of us (G.N.R.).

REFERENCE

DODD, C., and HU PAK MI, 1949, *Proc. Phys. Soc. B*, **62**, 154.

LETTERS TO THE EDITOR

The Dielectric Strength of Aluminium Oxide Films

Fröhlich (1937) has shown that, under the influence of an applied electric field F , electrons in a crystal will gain energy A from the field and lose energy B to the lattice vibrations, and that the resultant energy change $B - A$ is given by $B - A = B - (e^2/m)\tau F^2$, where e and m are the electronic charge and mass respectively, and τ is the average time between collisions. Electrical breakdown occurs when electrons possessing energy greater than the ionization energy J gain energy from the field. Accordingly, the breakdown field strength is inversely proportional to the square root of the relaxation time. Hence $F \propto l^{-1/2}$, where l is the mean free path of the electrons. If the latter is decreased by introducing foreign ions into the crystal, or by raising the temperature, the dielectric strength increases. In a thin film the value of B increases as the film thickness approaches the mean free path, since the number of collisions with the film boundary increases, and thus the dielectric strength of thin films of dielectric should rise with decreasing film thickness. Fröhlich has

calculated the order of such an increase for mica. Austen and Whitehead (1940) using mica of thickness varying from 10^{-4} to 10^{-5} cm. have verified his predictions, and more recently Plessner (1948) working with CaF, NaF and KBr films has also confirmed Fröhlich's theory.

This theory may be extended to films of amorphous materials, such as aluminium oxide, for which there should be a similar increase of strength although, since the electronic mean free path in amorphous substances is considerably less than in crystalline material, the effect should only occur to any marked extent in very thin films, for example films of thickness probably less than 100 Å.

The aluminium oxide films used in the present investigation were made by partially anodizing a thick aluminium film, formed by evaporation of the metal on an optical flat in a vacuum of 10^{-6} mm. Hg. The metal film was partially immersed in an aqueous solution of ammonium phosphate, and to prevent 'peeling' of the film the flat was subjected to ionic bombardment before evaporation.

The thickness of the oxide film was measured by employing equal chromatic order interference fringes. A cross section of the film is shown in Figure 1. It projects beyond the adjacent aluminium film, so that the oxide film may be defined by the two distances

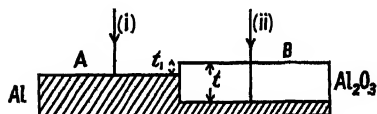


Figure 1.

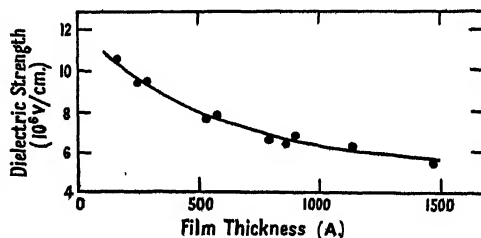


Figure 2.

t its thickness and t_1 the step between surfaces A and B. Hass (1949) has shown recently that the ratio t_1/t is constant, independent of t and equal to 0.27. Thus the optical path difference between rays (i) and (ii) reflected from the air-aluminium and the oxide-aluminium surfaces, respectively, is given by $2(\mu - 0.27)t$. Taking $\mu = 1.65$ —as given by Hass—the optical path difference between the rays is $2.76t$. When placed in contact with a semi-silvered optical flat, the step may be detected by means of fringes of equal chromatic order, formed by reflection at the film-optical flat surfaces, white light and a constant deviation spectrometer being used to produce these fringes. Since a thin surface oxide layer exists at the air-aluminium interface, phase changes at the oxide-aluminium and air-aluminium interfaces are equal. The fringes on both sides of the step were parallel and, accordingly, the oxide films were of uniform thickness.

Attempts were made to measure the film thickness directly by completely anodizing aluminium deposited on a film base of gold, nickel or platinum. With this arrangement the film thickness could be measured independently of the refractive index and the constant t_1/t but such metals 'peeled off' the flat during electrolysis.

The potential difference across the oxide film was applied by means of two spherical electrodes—steel ball bearings $\frac{1}{8}$ in. in diameter—one in contact with the surface A and the other lightly sprung so as to rest on B. The pre-breakdown currents, measured with gradually increasing potentials, were similar to those obtained by von Hippel (1938) with mica and alkali halide crystals. They are probably produced by strong field emission from the cathode although it has been suggested that they may be formed by ejection of electrons from impurity centres at high field strengths.

At a certain voltage this pre-breakdown current suddenly increased to a very high value and this voltage represents the breakdown value. Many measurements were made with the same film at different places on the film and averaged. In all experiments the film was in a vacuum and at 15° C. Figure 2 shows the variation of dielectric strength with film thickness. As the latter increased from 100 Å. to 1,000 Å. the strength decreased from 11×10^6 v/cm. to 6×10^6 v/cm. The value at 1,000 Å. agrees with that found by Güntherschulze and Betz (1936) for aluminium oxide in bulk.

The dielectric strength of one film was measured at 15° C. and at 100° K. It increased from 8×10^6 v/cm. to 9.5×10^6 v/cm. Similar results were found by von Hippel and Maurer (1941) using soda lime glass. This indicates that aluminium oxide has an amorphous structure and supports Hass' statement that no pores exist in the oxide film formed by anodization.

The relationship between formation voltage and film thickness in the present experiments is similar to that found by Hass (1949) and by Güntherschulze and Betz (1936). It is linear, but whereas the latter obtained a slope of 11 A/volt and Hass 13 A/volt, the present work gives 14 A/volt.

The Physical Laboratories,
The Washington Singer Laboratories,
University College, Exeter.
7th July 1950.

P. D. LOMER.

AUSTEN, A. E. W., and WHITEHEAD, S., 1940, *Proc. Roy. Soc. A*, **176**, 33.
FRÖHLICH, H., 1937, *Proc. Roy. Soc. A*, **160**, 230.
GÜNTHERSCHULZE, A., and BETZ, H., 1936, *Z. Phys.*, **100**, 539.
HASS, G., 1949, *J. Opt. Soc. Amer.*, **39**, 532.
VON HIPPEL, A., 1938, *Phys. Rev.*, **54**, 1096.
VON HIPPEL, A., and MAURER, R. J., 1941, *Phys. Rev.*, **59**, 820.
PLESSNER, K. W., 1948, *Proc. Phys. Soc.*, **60**, 243.

Ferromagnetic Resonance in Manganese Arsenide

Manganese arsenide is known to have a ferromagnetic critical point in the region of 40° C. and to exhibit pronounced temperature hysteresis (Bates 1928, 1929, 1931, 1933). Below the critical point ferromagnetic properties are clearly marked, and at room temperature the material has a saturation intensity of magnetization comparable with that in nickel at the same temperature. An investigation has been made into the ferromagnetic resonance absorption in this material at a wavelength of 1.25 cm. and at temperatures from 18 to 50° C.

The intensity of magnetization in small rod samples was measured by a simple ballistic method throughout the required temperature range and in applied magnetic fields up to 5,000 oersteds. It will be seen that at room temperature a value of about 400 E.M.U. is found for I , whereas Bates (1933) obtained a value of about 600; in each case satisfactory checks against a pure nickel sample were made. The reason for this discrepancy is being investigated.

From each rod, very thin discs were cut and mounted to form part of the lower end of a cylindrical cavity resonating in an H_{111} mode. The power absorbed in the specimen, at constant microwave frequency, was measured as a function of the strength of an applied steady magnetic field. The applied field, H_{\max} , for maximum power absorption was thus determined at each temperature. Using the relations first developed by Kittel (1947, 1948), it follows for a specimen of the form used that

$$g = \frac{2 \times 10^{-7}}{\lambda_s (BH)_{\max}^{1/2}},$$

where g is the Landé splitting factor, λ_s is the free space wavelength in centimetres, and B and H are expressed in kilo-oersteds.

Specimen results obtained with one sample are shown in the Table. Near the critical point the intensity of magnetization is changing very rapidly with temperature, the total absorption is small, and the experimental accuracy is therefore decreased.

	$\lambda_s = 1.25_2$ cm.										
t (° C.)	20.0	29.8	33.1	35.0	36.0	37.0	37.8	39.0	39.75	41.7	45 and 50
H_{\max} (oe.)	2700	2725	2800	2800	2925	3050	3280	3650	4150	~4550	4500 to 5000
I (E.M.U.)	408	390	383	378	365	331	267	120	50	—	—
g	3.7 ₂	3.7 ₈	3.7 ₁	3.7 ₂	3.6 ₈	3.6 ₈	3.6 ₈	3.9 ₈	3.8 ₄	3.7 ₈	3.4 ₄ to 3.8 ₀

Three different specimens were examined and, within the limits of experimental error, the g -value determined by the above equation does not vary up to the critical point. At room temperature, the three g -values found were 3.7_2 , 3.7_5 , and 3.8_2 .

It is of interest that resonance absorption could be detected above 45°C . where the manganese arsenide has paramagnetic properties. The absorption is small and broad, so that the position of the peak can only be found approximately, but it appears to be such as to suggest little, if any, change from the ferromagnetic g -value. There is no discontinuity observable at the critical point, the width of the absorption curve slowly increasing with the temperature. At 50°C ., the width is about 20% greater than at room temperature. These latter observations are in accordance with those of Bloembergen (1950) whose results for nickel and supermalloy were published during the course of this investigation.

Of particular interest, however, is the very large g -value which is obtained. Previously published results by various authors (see Kittel 1949) have given values mainly of the order of 2.20 and the excess of this above the value 2.00 for an electron spin has been qualitatively interpreted in terms of the contribution of the orbital moment (Kittel 1949, Polder 1949, Van Vleck 1950). Kittel has shown that if g is the g -value deduced from microwave absorption measurements and g' that from direct gyromagnetic measurements, g should be greater than 2.00 by the amount that g' is less than 2.00 for small variations about the electron spin value.

If this type of explanation is to hold in the case of manganese arsenide, one would expect g' to be very much less than 2.00. It is thought that there has been no such measurement on manganese arsenide itself, but Galavics (1939) has reported measurements on an Mn-Sb alloy stated to be Mn_3Sb , for which he obtained a value for g' between 1.9 and 2.0.

It is hoped to continue this investigation with other materials of the types Mn-P, Mn-Sb, Mn-Bi, and results will be reported later.

The authors wish to record their thanks to Professor L. F. Bates, who kindly supplied the manganese arsenide samples, for his continued interest and advice.

The University,
Nottingham.
25th July 1950.

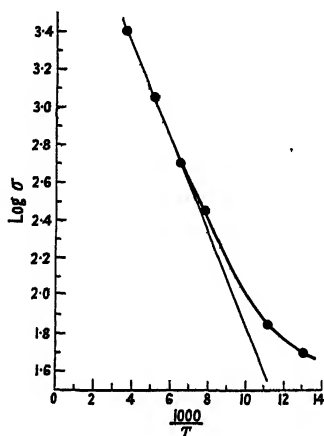
G. D. ADAM.
K. J. STANDLEY.

- BATES, L. F., 1928, *Proc. Roy. Soc. A*, 117, 680; 1929, *Phil. Mag.*, 8, 714; 1931, *Proc. Phys. Soc.*, 43, 87; 1933, *Phil. Mag.*, 16, 657.
BLOEMBERGEN, N., 1950, *Phys. Rev.*, 78, 572.
GALAVICS, F., 1939, *Helv. Phys. Acta*, 12, 581.
KITTEL, C., 1947, *Phys. Rev.*, 71, 270; 1948, *Ibid.*, 73, 155; 1949, *Ibid.*, 76, 743.
POLDER, D., 1949, *Phil. Mag.*, 40, 99.
VAN VLECK, J. H., 1950, *Phys. Rev.*, 78, 266.

Electrical Conductivity of Gray Tin

Gray tin has a non-metallic crystal structure similar to diamond (Bijl 1919, Brownlee 1950), i.e. it is the end-member of the series carbon-silicon-germanium-gray tin. One would therefore expect its electrical properties to be appropriate to its position in the series; but there seem to be no reliable experimental results available in the literature. It was originally reported that gray tin was a superconductor, but de Haas (1928) and Sharvin (1945) showed that this was incorrect, at least above 1.2°K . Busch (private communication) has obtained some values for electrical conductivity of gray tin powder which show it to be a semiconductor. On the other hand Moesveld (1937) reported that it was a normal metallic conductor. It seemed unlikely that results of real value could be obtained until a way was found of preparing gray tin in massive crystalline form free from white tin. Normally it is prepared by keeping tin at a temperature well below its transition point (13.2°C .) when the metal crumbles into a friable powder. The large change in specific gravity from 7.31 to 5.75 is no doubt the cause of this disintegration. Unfortunately, subsequent compression of the powder tends to transform it back to white tin, even when the compression is carried out at low temperature.

In transforming a kilogram of spectroscopically pure tin it was found that amongst the large quantity of fine powder produced were a few lumps of gray tin several millimetres in linear dimensions. Suitable electrodes were pressed against these lumps and their electrical conductivity measured for small D.C. currents, using a milliammeter and potentiometer. No contact non-linearity was observed, but precautions were necessary to eliminate the rather large thermo-electric E.M.F. generated under non-uniform temperature conditions.



α-tin: variation of conductivity with temperature.

The variation of conductivity with temperature is shown in the Figure; the values given for specific conductivity must be regarded as only approximate as there was some difficulty in measuring the cross section and electrode areas of the specimens. Almost identical results were obtained for three different specimens.

It is seen that gray tin is an intrinsic semiconductor, in which the conductivity is given by

$$\sigma = A \exp(-\epsilon/2kT).$$

From the upper portion of the curve we obtain $A \simeq 6 \times 10^4 \text{ ohm}^{-1} \text{ cm}^{-1}$, and $\epsilon = 0.098 \text{ ev}$. At the lower temperatures the presence of impurities in the tin is presumably causing enhanced conductivity and a lower activation energy. Gray tin thus fits consistently into the series of intrinsic homopolar semiconductors as shown in the Table below. The very small energy difference between the filled and unfilled bands is noteworthy, and leads to the high conductivity at 0° C ., remarkable for a semiconductor, of approximately $2.5 \times 10^8 \text{ ohm}^{-1} \text{ cm}^{-1}$.

	ϵ	A
Silicon	1.12	0.9×10^4
Germanium	0.75	3.3×10^4
Gray tin	0.1	6×10^4

The author wishes to acknowledge the award of a Leverhulme Research Fellowship.

Department of Electrical Engineering,
Imperial College of Science and Technology,
London, S.W.7.
6th July 1950.

J. T. KENDALL.

- BIJL, A. J., 1919, *Proc. Acad. Sci. Amst.*, **21**, 405.
BROWNLEE, L. D., 1950, *Nature, Lond.*, (in the press).
DE HAAS, W. J., 1928, *Proc. Acad. Sci. Amst.*, **31**, 350.
MOESVELD, A. L., 1937, *Z. Phys. Chem. A*, **178**, 455.
SHARVIN, G., 1945, *J. Phys., USSR*, **9**, 350.

Development of the Low Pressure Electrodeless Discharge in a High-Frequency Electric Field

The starting of a high-frequency electrodeless discharge with the mean free paths of electrons much larger than the dimensions of the glass vessel has recently been investigated (Gill and von Engel 1947, 1948). In these experiments a cylindrical glass tube containing gas at this low pressure was placed in a uniform high-frequency electric field (wavelength order 10 m.) directed parallel to the axis. It was shown that, if an electron starting at one end wall with a small energy and in a suitable phase of the field crossed the tube in one half-cycle, and hit the opposite end wall fast enough to release more than one secondary electron, then an electron avalanche would develop. The starting field strength should therefore depend on the material of the end walls, and not on the gas. This was confirmed by experiment. As the discharge develops a current passes through the gas and the field decreases. This working field strength, and the spectrum emitted, are characteristic of the gas, and do not depend on the material of the vessel.

It was derived that in the initial stages of the discharge electrons have maximum energy (about 100 ev.) when they hit an end wall. This is enough to release between one and two secondary electrons per incident electron, most of which contribute to the development of the avalanche, whilst a few are lost to the side walls. This maximum energy is also much greater than the ionization energy of any gas.

But in the fully developed discharge both photographic measurements at wavelengths of about 3,000 Å. and visual observation show that most of the light is emitted from the centre of the cylindrical vessel and very little from the ends. Assuming that this light comes directly from atoms excited by collisions with electrons, it appears that the electrons are fastest in the central region. A cloud of electrons seems here to oscillate with small amplitude about a space charge of positive ions in the centre of the tube. Very few of these electrons hit the end wall; if they did they would be lost. But it is known that when losses to the walls are large (e.g. beyond cut-off wavelength (Gill and von Engel 1948) no field, however large, can start the discharge.

As far as we know no attempt has been made to describe the transition from the initial to the final stage of such a discharge. At first sight it would appear that, of the few ions formed in the gas in the earlier stages, most would be produced near the walls where the electrons are fastest. Two clouds of positive space charges would then be formed near the end walls, and this would continue to make the electrons fastest near the ends of the tube. (The effect of the positive wall charge caused by secondary electron emission has been considered, but found irrelevant here.)

This, however, is not true. A calculation of the energy of an electron as a function of position along the tube shows that any electron capable of starting the discharge acquires energy equal to the ionization energy of any gas well before it reaches the centre of the tube. The ionization produced is obtained by combining this result with known ionization probabilities.

Now electrons crossing the tube in one direction produce a distribution of ions in space which increases rapidly towards the centre, and then slowly from the centre to the far wall. Electrons crossing in the opposite direction produce a symmetrically opposite distribution. The sum gives the maximum number of ions at the centre. As the multiplication process continues the number of ions increases, but the maximum ionization is always at the centre.

When the number of ions becomes considerable the field in which electrons move is the applied field plus the field due to this positive space charge. As the latter increases the electrons reach their maximum speed progressively nearer the centre of the tube. New electrons are now produced (by ionization of gas atoms) near the centre, and start oscillating in a phase different from that of the main cloud. Some are lost to the walls; most oscillate in the gas with a maximum speed near the centre, their amplitudes becoming smaller the larger the positive ion space charge. These are the electrons which are found in the final stage of the discharge. The electrons originally produced by secondary emission from the walls are progressively lost to the end walls as their speed at the ends

of the tube is reduced by the influence of the positive ion space charge. These electrons neutralize the positive wall charge.

Positive ions in the gas drift to the walls owing to self-repulsion in the ion cloud. However, electrons produced in the gas in unfavourable phases of the applied field are also lost to the wall, neutralizing these positive ions and charging the walls slightly negatively.

Thus it appears that the transition from multiplication by secondary emission to an equilibrium state sustained only by ionization in the gas can only occur when the electrons produced by the first process are all replaced by electrons produced in the gas which then oscillate with an amplitude smaller than the length of the vessel.

The Clarendon Laboratory,
Oxford.
10th August 1950.

G. FRANCIS.
A. VON ENGEL.

GILL, E. W. B., and VON ENGEL, A., 1947, *Nature, Lond.*, 159, 404; 1948, *Proc. Roy. Soc. A*, 192, 446.

REVIEWS OF BOOKS

An Introduction to Heat Transfer, by M. FISHENDEN and O. A. SAUNDERS. Pp. x+205. 1st Edition. (Oxford: The University Press, 1950.) 15s.

There seems to be a general trend towards putting the engineering treatment of heat-transfer problems on to a more systematic basis than has been the practice in the past, and the authors of this book have been among those who were instrumental in bringing it about.

They have now published a book, distinct from their earlier joint writings, in which the whole subject is surveyed on a reasoned basis.

The first chapter deals with radiation, whether the theoretical simple radiation from a black body or the complicated phenomena encountered, for example, when considering the radiation from the flames in a furnace. The questions which arise when geometrical considerations intervene, or when the emitting and receiving surfaces are of different emissivities, are considered.

The next chapter is on conduction, and does not deal in detail with the calculation of heat flow in complex systems, though it does treat the theory of sufficient cases for most practical applications. The theory is a matter for mathematicians, and is treated in separate books, of which that by Carslaw and Jaeger is probably the best known, though a recent one by Ingersoll, Zobel and Ingersoll should not be overlooked, especially as it and the book under review are in a sense complementary, and share the same outlook—that of seeing basic physics in view while applying the theory in practice. In this chapter, Fishenden and Saunders describe in detail the graphical method due to Schmid for calculating heat flow, and they also refer to the use of relaxation methods, and to electrical analogies.

About 60% of the book is devoted to convection, first the general theory, then the relation between convection and friction, then forced and natural convection in the commonest shapes of system, concluding with convection to and from fluids which evaporate or condense in the course of the action.

The material is admirably presented, and is illustrated by worked examples, which add considerably to the value of the exposition. Many useful charts and tables are given, which will ease the work of calculation in specific applications to real systems.

J. H. A.

Die Übmikroskopie, by B. VON BORRIES. Pp. 416, with 70 plates and 225 line diagrams. 1st Edition. (Aulendorf, Württemberg: Editio Cantor, 1949). 48 D. Marks.

For some ten years now there have been available commercial models of the electron microscope which can far surpass the optical microscope in resolving power and which, although several times as expensive as the latter, are not exorbitantly priced as modern physical apparatus goes. Nevertheless the amount of research work of fundamental importance with which it can be credited remains relatively small. After considerable

enthusiasm, in principle, over the introduction of the new instrument, there has followed in many quarters a certain caution in regard to its practical application. This degree of disillusionment has been due in some part to a lack of understanding of the natural limitations attaching to its technique, so that an immediate extrapolation was often expected to be possible from optical practice into the smaller orders of dimensions. In fact the preparative methods, and the way in which image contrast arises, are so different in electron microscopy that quite new interpretative methods are required. This is not to say that a great deal of important progress has not already been made in metallurgy, biology and most of the applied sciences with the aid of the electron microscope. But much more might have been possible if a clearer realization had existed of the natural limitations to its use.

It is with the exploration of these limitations that Dr. von Borries, who was responsible with Dr. E. Ruska for the original development of the magnetic electron microscope, has principally concerned himself in recent years. An account of these researches forms the core of his present book. He has investigated the instrumental factors of illumination brightness, which determines the maximum magnification which can be used and hence the resolution limit, the recording sensitivity, which sets a lower limit to the total amount of energy passing through the specimen during the exposure necessary to form an image, and the fluorescent efficiency of viewing screens, which determines the lowest permissible illumination for focusing and hence the temperature rise in the specimen. In the light of these findings he examines the possibility of electron microscopy of living material, as limited by the thermal and ionizing effects of the electron beam. The lethal dosage is so quickly reached that observation would only be practicable at low magnification, and hence at low resolving power; only in the case of a few organisms could the optical limit be exceeded, even if a sufficiently high voltage were available to penetrate them.

Aside from the biological aspect of the enquiry, with its pessimistic conclusion, the author presents a great deal of valuable information as to the efficiency of electron guns, the scattering and energy loss which accompany the passage of electrons through matter, and the characteristics of photographic emulsions towards electron illumination. The present limit set to resolving power by the aberrations of electrons is also fully discussed, although possible means of correction are hardly mentioned.

As the book is primarily intended for scientists who may wish to assess the positive value of the electron microscope in their particular field of research, the detailed exposition of its limitations is preceded by a review of its principles and constructional features, and is followed by a survey of its applications in pure and applied science, the three sections being of approximately equal length. An appendix has been added to bring the treatment up to date to the beginning of 1949, the main text having been completed a year earlier. The result admirably fulfils the author's intention of giving a well-balanced account of the present state of development of electron microscopy. Omissions of reference to non-German work, which in the difficult war and post-war conditions of communication might have been expected to be serious, are relatively minor, although it is admitted that in many cases the author has not had access to the full text of a paper cited. The illustrations are naturally almost entirely from German sources, and thus permit a much better assessment of the progress made in electron microscopy there since 1939 than was possible from the compressed and unillustrated FIAT reports.

The quality of the reproduction is high, but the same cannot be said of the rest of the technical production of the book. It is probably inevitable that the quality of the paper for the text should be poor, in present-day circumstances, but the binding could surely have been better executed. The pages have been inserted individually, not in stitched and folded sections, so that they come loose with very slight use. It is regrettable to have to make such a complaint as to the production of what is in all other respects an excellent monograph.

V. E. COSSLETT.

Colours and How We See Them, by H. HARTRIDGE. (Royal Institution Christmas Lectures 1946-47). Pp. xi + 158. 1st Edition. (London: G. Bell & Sons Ltd., 1949). 15s.

This book is based on the Royal Institution Christmas Lectures which Professor Hartridge gave to boys and girls in 1946, and the spirit of entertainment and instruction always associated with these lectures is well reflected in its pages. It is essentially an

experimental book in which a great many phenomena are mentioned. These include the spectrum, spectral absorption, colour mixture, interference, diffraction, applications of colour in science and medicine, dichroism, the effect of intermittent illumination and so on. Anyone invited to give a lecture on colour will find a wealth of demonstrations from which to draw, many of the experiments being illustrated in colour. The theoretical explanations of the phenomena do not go very deeply into the subject, but that is hardly to be expected in a book of this type. Quite reasonably, Professor Hartridge has not considered this the appropriate opportunity to enlarge on his theoretical ideas about colour vision.

W. D. WRIGHT.

Cathode-Ray Tube Traces, by HILARY MOSS. Pp. 66. 1st Edition. (London: Electronic Engineering, 1949.) 10s.

This monograph, based on articles published by the author in the journal *Electronic Engineering*, should be of interest to all users of the cathode-ray tube.

The subject matter in general is well presented and the mathematical arguments are not unduly complicated.

The first chapter merits separate mention, in that it deals in detail with the various forms of Lissajous Figure and, in addition to the basic theoretical considerations, points out various practical pitfalls.

Chapter II is disappointing, both in title ('Straight Line Time-Bases') and in content. It is surprising that the linear sawtooth waveform should be dismissed in a mere section of this chapter when various other time-bases of comparatively rare occurrence, such as the circular and spiral forms, should merit the whole of Chapter III.

On reading these chapters one gathers, perhaps wrongly, that Dr. Moss has over-concentrated his attention on time-bases of sinusoidal origin rather than on those of discontinuous form.

The remainder of the book is mainly devoted to the interpretation of various waveforms more usually encountered as the phenomenon under investigation.

A résumé is given of Fourier series and application is made to several of the common pulse waveforms. Beats and the amplitude modulated envelope are also treated.

The author deserves congratulations for his illustrations, but the bibliography is scanty and would only serve as a first guide to further reading.

A. V. LORD.

Photoelasticity: Principles and Methods, by H. T. JESSOP and F. C. HARRIS. Pp. viii + 184. 1st Edition. (London: Cleaver-Hume Press Ltd., 1949.) 28s.

A book on photoelasticity emanating from the Laboratory whence came the classic treatise on the subject by Coker and Filon commands immediate interest and attention. Work on photoelasticity continues at University College, London, and the senior author was the first Chairman of the recently formed Stress Analysis Group of the Institute of Physics.

Authors are not responsible for their publisher's advertisements, but the assertion on the dust jacket that it is a reference work is supported by the authors' statement in the preface that they have given in the text the sources of any original work quoted. Unfortunately, and this is the most serious criticism of the book, it is not true. Statements are often made without the name of the worker responsible being given, seldom if ever is a reference accompanied by chapter and verse, and there is not even a bibliography from which one might hope to guess at the origin. The tabulated data in the appendix would inspire more confidence if reference were made to the sources.

A laboratory such as that of the authors which is the centre of so much work on the subject provides an *embarras de richesses* and the authors seem to have had difficulty in selecting and arranging their material. The book gives the impression of having been hurriedly put together, and not enough care has been spent in critically examining it before printing. For example on page 39 the authors state that a convenient convention of signs is the Cartesian; yet on page 43, they give the distances from a principal point of a lens system to the corresponding principal focus, H_1F_1 or H_2F_2 , as both equal to f_0 ; whereas in

a Cartesian system one must be the negative of the other. Again on page 63, the definitions of the stress-optical coefficients leave the reader guessing as to which refers to the ordinary and which to the extraordinary ray, and what is the sign of the difference between them. Again on page 60 a compensator is said to be between 'crossed' polarizer and analyser, yet it is not until page 65 that a definition of this term is given. When much is left vague, the exact analysis of the mode of working of quarter-wave plates given in the appendix, seems a little out of place, especially in a book which sets out to be a guide to "those who wish to undertake practical work on stress-exploration".

The book is well printed and the diagrams are clear, but the half tone illustrations are poorly reproduced; it is a pity these were not grouped together and printed separately on glazed paper; justice would then have been done to the skill of Mr. Storrer the photographer. It would seem that the worked examples are invaluable, and the descriptions of the practical requirements of the optical bench are admirable, although emphasis should have been placed on the need for examination in parallel light, more particularly when the specimens are thick and highly stressed.

It is to be hoped that when the authors prepare a second edition they will expand Chapters IV and V on polarization and photoelasticity, and will give references throughout, so as to enable those who may wish to go more deeply into the subject to do so with less difficulty.

A. M. TAYLOR.

Symposium on Electronics, edited by A. G. PEACOCK. Pp. xiii + 199. 1st Edition. (London: Chapman and Hall in collaboration with S.I.M.A., 1949.) 16s.

The first of a series of annual Symposia on Electronics in Research and Industry, organized by the Scientific Instruments Manufacturers Association, was held in London in 1948. The papers presented at this Symposium are published here in book form, and cover a variety of applications of electronic methods to the solution of instrument problems. Recent advances in the measurement of frequency, small displacements, ionizing radiations, low gas pressures and sound are described, and accounts are also given of the use of electronic techniques in computing, the radiosonde, ultrasonics, spectroscopy, picture telegraphy and metal detection. Because of the large number of subjects covered the treatment of each is brief and descriptive. For this reason the book is primarily of interest only to those non-specialists who wish to have a general summary of a few of the many ways in which electronic equipment is being applied, both in research and in industrial production. Accounts are given of the discussions which followed the presentation of each paper, but the space so occupied could, in the reviewer's opinion, have been more usefully allocated to the expansion of some of the papers.

J. M. MEEK.

Ultra-Violet and Daylight Rays, by J. R. ASHWORTH. Pp. x + 73. 1st Edition. (Liverpool: University Press, 1949.) 6s.

This monograph presents the results of Dr. Ashworth's measurements of ultra-violet and visible light received at ground level over the 16 years from 1932 to 1947. He has used a simple type of photographic measuring technique, which is described and compared with other techniques. The results are given in the form of tables and graphs, showing the variation of the average intensities of visible and of ultra-violet light, with wavelength near 3600 Å. measured in arbitrary units. Evidence is given that these very simple measurements have some real significance, and are not unduly affected by the characteristics of the particular place where they were made (Rochdale).

The results are interpreted as showing that the visible light emitted from the sun is roughly constant throughout the sunspot cycle, but that the ultra-violet is more intense at times of sunspot maximum. These conclusions are, however, not directly deducible from the observations because, according to the author's theory, the variable ionization produced by the ultra-violet in the upper atmosphere in turn produces variable absorption in both the ultra-violet and the visible light reaching the earth.

It is unfortunate that the theory is left in a very speculative and indefinite form. It is suggested that the absorption occurs in Region E of the ionosphere, and is due somehow to the ionization of this region; it would be interesting to know in more detail what Dr. Ashworth believes this mechanism of absorption to be. Reference to his published papers fails to provide any more detail.

J. A. RATCLIFFE.

Les ondes électromagnétiques centimétriques, par LOUIS DE BROGLIE. Pp. 274. (Paris: Éditions de la 'Revue d'Optique Théorique et Instrumentale', 1948.) 800 fr.

Each year since 1944 M. Louis de Broglie has organized a conference to summarize the state of knowledge in some important branch of physics. In 1947 one of these conferences, occupying five days, was held to discuss what was known of centimetric electromagnetic waves, and the proceedings are recorded in the volume here reviewed.

The object was not to repeat the numerous accounts which have already appeared of the workings of radar, and of the magnetron, which led to the possibility of centimetre radar, but to choose nine separate and less well-known aspects of the subject and to ask nine separate authors to report on these. These reports are collected in this volume. They deal with the subjects of: the exchange of energy in electron tubes; the effects of space charge in velocity-modulated tubes; the effects of curvature in waveguides; excitation of and coupling to cavity resonators; measurements at centimetre wavelengths; aerials; propagation; radiation from the sun and the galaxy; the use of centimetre waves in nuclear physics.

In a conference of this kind it was inevitable that the treatment should differ from subject to subject. Some of the papers describe some special point rather fully; those on space charge and on curved wave guides are of this type. Others, for example those on aerials and propagation, attempt to provide a broad survey without much detail.

All the subjects dealt with are advancing so rapidly that a report of a conference held three years ago must inevitably seem somewhat out of date. Since the study of extra-terrestrial radiation was only started seriously after the end of the war, it is perhaps only natural that the corresponding section of the report now seems most out of date. It is also a little surprising to see this subject, whose advance has depended so largely on measurements of metre waves, so fully discussed in a volume dealing with centimetre waves.

J. A. RATCLIFFE.

Recent Advances in the Physiology of Vision, by H. HARTRIDGE. Pp. xii+401. (London: J. and A. Churchill Ltd., 1950.) 25s.

Much of Professor Hartridge's book is concerned with topics belonging to physiological optics and colour physics, in which physiologists and physicists have been equally active. Eye motions and the dioptric properties of the eye, visual resolving power, spectral luminosity functions, the directional sensitivity of the retina, quantum notions in vision, colour discrimination, the trichromatic and other colour theories, etc., are treated, along with the topography of the nerve fibres in optic nerve, optic tract and visual cortex, the electrical response of the retina, the effect of light stimuli on the Berger rhythms, the visual pigments, electric and magnetic phosphenes, and so on. The author's object has been to survey a very wide field, giving enough of the older work to make clear where advances have been made.

It is probable that no worker in the subject of vision will read this book without wanting to take issue with Professor Hartridge on some of his contentions. But a work of this kind, particularly one dealing with vision, can scarcely be criticized for not presenting a coherent and generally accepted picture. It can be judged successful if it draws the reader's attention to new results and gives some idea of how people are interpreting or speculating about them. Professor Hartridge's book does all these things, and is, perhaps, particularly strong on the 'speculating'. The style is easy, but the arrangement of material leaves something to be desired. In a number of cases the same subject is dealt with in widely separated sections. For example, eye-motions and their measurement are discussed in Section 8 (p. 64) and in Section 54 (p. 329), but there seems no justification for dividing the material in this way. Among several topics omitted which might have been expected to earn a place may be

mentioned particularly W. J. Schmidt's polarized-light analysis of the structure of the outer segments of the retinal rods, an analysis largely confirmed by the electron microscope photographs of rods recently obtained by Sjöstrand.

Of the author's handling of visual theories it is sufficient to say that most physicists would demand more precise and quantitative reasoning to support some of the statements made. A few comments on other aspects of the presentation may be added. Although the crux of Hecht, Shlaer and Pirenne's work on quanta and visual thresholds is the relation between the 'sharpness' of the threshold and the minimum number of quanta involved, no mention is made of this relation in the two pages devoted to the work in question. The results of Mandelbaum and Mintz on the dark adaptation of the colour receptors, briefly described on p. 50, are stated to refer to rod vision. This is clearly a mistake, which is aggravated by a suggested connection of the results with dark-adaptation measurements by Miles which certainly do refer to rod vision. In deploring the lack of information on the variation of visual acuity with pupil size the author seems to have overlooked the extensive measurements of Fabry and Arnulf published in 1937 (*La vision dans les instruments*, Editions de la Revue d'Optique). An explanation is given of how the binocular threshold may be lower than the monocular threshold purely on account of 'probability summation' (pp. 237-238). This is immediately followed by a statement implying that any such difference must arise from physiological summation. It is asserted (p. 203) that the coefficient curves for protanomalous subjects show no negative coefficients in any part of the spectrum when primaries 650, 530, and 460 $m\mu$ are used, and the curves shown in Figure 160 are cited in support. The figure is obviously wrong, as at wavelengths below 460 $m\mu$ the coefficients do not add up to unity, and the statement itself is incorrect.

A number of other criticisms of the same kind could be made, but taken as a whole the book provides a collection of facts and current speculations on vision which will prove of value to all students of the subject.

W. S. S.

Méthodes de calcul dans des problèmes de mécanique. Pp. 102. (Paris: Centre National de la Recherche Scientifique, 1949.) No price.

In the spring of 1948, two international conferences on methods of calculation in problems of mechanics were organized at Marseilles and Paris by the C.N.R.S.; the papers contributed to these conferences have been collated and published in this volume. The contributors and topics are as follows: J. Valensi (Marseilles), "The role of applied mathematics in engineering"; D. N. de G. Allen (London), "Relaxation methods and problems of frameworks"; "Relaxation methods and the solution of differential equations", together with "Supplementary notes on the application of relaxation methods"; Th. Vogel (Marseilles), "The escalator method for the calculation of eigenfrequencies"; M. Picone (Rome), "New points of view in harmonic analysis"; L. Couffignal (Paris), "Calculating machines for harmonic analysis according to the method of H. and Y. Labrouste"; J. M. Burgers (Delft), "Problems related to the theory of turbulence"; L. Malavard (Paris), "Some recent applications of the method of electrical analogies"; A. van Vijnngaarden (Amsterdam), "Potential flow about a solid of revolution"; F. H. van den Dungen (Brussels), "The application of the calculus of variations in fluid mechanics"; L. Couffignal "The role of numerical calculation in scientific and technical research".

The papers naturally vary in style and standard but they are, without exception, lucidly written and the book can be thoroughly recommended to a physicist or engineer who wishes to acquire some idea of modern techniques used in the realm of classical mechanics. R. M. D.

Glass-to-Metal Seals, by J. H. PARTRIDGE. Pp. xii+238. 1st Edition. (Sheffield: Society of Glass Technology, 1949.) 25s.

This is a comprehensive and authoritative guide to the art of sealing metals to glasses. The production of graded glass-to-glass seals is therefore also dealt with in detail. The problems of stress distribution, so vital in determining success or failure, are ably and adequately discussed.

This book should be available to all workers, both scientific and technical, in every physics and chemistry laboratory; it is first-class.

G. I. F.

CONTENTS FOR SECTION A

	PAGE
Mr. B. D. HYAMS, Mr. M. G. MYLROI, Mr. B. G. OWEN and Dr. J. G. WILSON. A Magnetic Cosmic-Ray Spectrograph with Counter Recording	1053
Mr. B. G. OWEN. A Magnetic Cosmic-Ray Spectrograph with Counter Recording—II: The Electronic and Recording System	1074
Dr. E. P. GEORGE and Mr. A. C. JASON. Observations on Cosmic-Ray Penetrating Showers at High Altitude, Sea Level and Below Ground	1081
Dr. D. M. RITSON. Slow Mesons in the Backward Flux of the Cosmic Radiation	1098
Prof. L. JÁNOSSY and Mr. H. MESSEL. Fluctuations of the Electron-Photon Cascade—Moments of the Distribution	1101
Mr. L. R. B. ELTON. The Effect of Nuclear Structure on the Elastic Scattering of Fast Electrons	1115
Mr. J. IRVING. Non-physical Solutions in Classical Finite Electron Theory	1125
Dr. A. LAGERQVIST, Mr. E. LIND and Dr. R. F. BARROW. The Band-Spectrum of Barium Oxide	1132
Dr. M. AFAF. Band-Spectrum of ZrO	1156
Letters to the Editor :	
Mr. F. K. GOWARD and Mr. J. J. WILKINS. The Photo-Disintegration of Oxygen into Four Alpha-Particles	1171
Mr. J. J. WILKINS and Mr. F. K. GOWARD. Ground State ^8Be Nuclei in Photo-Disintegration Stars	1173
Mr. F. C. W. COLMER and Mr. D. J. LITTLER. Pile Neutron Absorption Cross Sections of some of the Elements	1175
Mr. D. J. RAVENHALL. The Effect of Screening of the Cross Section for Pair Production by Electrons	1177
Dr. J. H. FREMLIN and Miss MADELINE C. WALTERS. Background Tracks in Electron-Sensitive Nuclear Emulsions	1178
Mr. J. DARBY, Dr. J. HATTON and Dr. B. V. ROLLIN. The Attainment of very Low Temperatures by a Two-Stage Adiabatic Demagnetization Process	1179
Mr. J. DARBY, Dr. J. HATTON and Dr. B. V. ROLLIN. Superconductivity of Lead Sulphide, Selenide and Telluride	1181
Dr. K. MENDELSSOHN and Mr. J. L. OLSEN. Heat Flow in Superconductive Alloys	1182
Mr. J. K. HULM. The Dielectric Properties of some Alkaline Earth Titanates at Low Temperatures	1184
Reviews of Books	1185
Contents for Section B	1186
Abstracts for Section B	1186

ABSTRACTS FOR SECTION A

A Magnetic Cosmic-Ray Spectrograph with Counter Recording, by B. D. HYAMS, M. G. MYLROI, B. G. OWEN and J. G. WILSON.

ABSTRACT. An instrument is described in which the momentum of single cosmic-ray particles is measured, the particles leaving the instrument in a sufficiently collimated beam for use in subsequent experiments. The 'maximum detectable momentum' of the apparatus is about 3×10^{10} ev/c., almost all particles are collected within 10° of the zenith and in directions individually defined to $\pm 1^\circ$, the output is about 400 particles/day, and the proportion of spurious records is of the order of, or less than, 0.1%.

A Magnetic Cosmic-Ray Spectrograph with Counter Recording—II: The Electronic and Recording System, by B. G. OWEN.

ABSTRACT. The paper describes the electronic selection and recording systems of the spectrograph and the performance and overall serviceability of the equipment as shown in continuous operation over about one year.

Observations on Cosmic-Ray Penetrating Showers at High Altitude, Sea Level and Below Ground, by E. P. GEORGE and A. C. JASON.

ABSTRACT. Penetrating showers have been investigated, using Geiger counters, at a depth of 30 m. below ground, at sea level and at an altitude of 3,457 m. The influence of the geometry of the absorbers was studied. The transition curves for local showers, corrected for this geometric effect, give the following collision lengths of the generating particles: Pb, 180 ± 40 ; Al, 85 ± 15 ; Paraffin ~ 80 gm/cm². These values correspond to a collision cross section close to the geometric cross section of the nuclei. From the results obtained for extensive penetrating showers, it is concluded that most of the penetrating particles in these showers appear to be generated in the atmosphere. In 40 days, no penetrating showers were recorded below ground. Using photographic plates, the density distribution of shower tracks associated with nuclear disintegrations was found to be similar at sea level and 3,457 m.

Slow Mesons in the Backward Flux of the Cosmic Radiation, by D. M. RITSON.

ABSTRACT. Using the method of 'delayed coincidences' measurements have been made at sea level on the flux of slow mesons travelling backwards from the ground.

Fluctuations of the Electron-Photon Cascade—Moments of the Distribution, by L. JÁNOSSY and H. MESSEL.

ABSTRACT. The second moments of the electron-photon cascades are evaluated. It is found that the mean square fluctuation at the cascade maximum is that expected for a Poisson distribution or less; before and after the maximum the fluctuation exceeds greatly that expected for a Poisson distribution. The correlation coefficient between electrons and photons is zero for small thicknesses or near the cascade maximum. For other regions the correlation coefficient is positive. Both mean square fluctuation and correlation between electrons and photons can be understood qualitatively in terms of fluctuation of effective depth.

The Effect of Nuclear Structure on the Elastic Scattering of Fast Electrons, by L. R. B. ELTON.

ABSTRACT. The cross section for the scattering of electrons by atomic nuclei is investigated at energies for which the nuclei can no longer be treated as point charges. Two simple nuclear models are used. A general expression is obtained using Born's approximation, and an exact numerical calculation is carried through for 20 mev. electrons scattered by gold nuclei. It is concluded that at this energy considerable deviations from the formulae which treat the nucleus as a point charge should be expected, and that these should furnish some information about the charge distribution within the nucleus.

Non-physical Solutions in Classical Finite Electron Theory, by J. IRVING.

ABSTRACT. An approximate linear form of the Peierls-McManus equations of motion for an electron is shown to yield runaway solutions for a particular influence function, which otherwise satisfies all the conditions imposed by the theory. The general problem is to discuss the existence of complex roots of a transcendental equation for an arbitrary form of the influence function. It is shown that this equation always has roots. These, however, may be such that the corresponding motion of the free electron is of a damped nature. It has not been found possible to construct a function which avoids runaway solutions or to prove whether or not such a function exists when all conditions are satisfied.

The Band-Spectrum of Barium Oxide, by A. LAGERQVIST, E. LIND and R. F. BARROW.

ABSTRACT. The band-spectrum of BaO between 5000 and 7000 Å., which consists of a single ${}^1\Sigma \rightarrow {}^1\Sigma$ system, has been the subject of a new rotational analysis in which the following bands have been examined: 5,0, 4,0, 3,0, 3,1, 2,0, 2,1, 1,1, 1,2, 0,2, 0,3 and 0,4. The constants derived are:

$$\begin{aligned} B_v'' &= 0.3124_9 - 0.0013_0(v'' + \tfrac{1}{2}) - 0.0000_2(v'' + \tfrac{1}{2})^2 & B_v' &= 0.2584 - 0.0011_1(v' + \tfrac{1}{2}) \\ D'' &= 26.5 \times 10^{-8} & D' &= 28 \times 10^{-8} \\ r_e'' &= 1.940 \times 10^{-8} \text{ cm.} & r_e' &= 2.133 \times 10^{-8} \text{ cm.} \\ \omega_0'' &= 669.8_1, & x_e''\omega_0'' &= 2.05_4 & \omega_0' &\sim 500, & x'\omega' &= 1.6 \\ & & \nu_{0,0} &= 16,722.2_5 \text{ cm}^{-1}. \end{aligned}$$

In the upper state some 16 perturbations have been found. They appear to arise from interactions with at least four electronic states or substates. Fairly complete information about one of the perturbing levels has been obtained. The constants of this state are:

$$\begin{aligned} B_v &= 0.2254 - 0.0013_5(v + \tfrac{1}{2}), \\ G_v &= 450.4_4(v + \tfrac{1}{2}) - 2.9_9(v + \tfrac{1}{2})^2, \\ \nu_{0,0} &= 17476.7 \text{ cm}^{-1}. \end{aligned}$$

The band-constants of the other perturbing states are very similar to those given above.

Band-Spectrum of ZrO, by M. AFAF.

ABSTRACT. The spectrum of ZrO has been photographed in the ultra-violet, the visible and the infra-red, from arcs running under special conditions. This has made possible the identification of three new systems in the ultra-violet, as well as two less conspicuous ones in the infra-red. It has also made it possible to extend the blue system α of ZrO.

THE PROCEEDINGS OF THE PHYSICAL SOCIETY

Section B

VOL. 63, PART 11

1 November 1950

No. 371 B

A Note on the Orientated Overgrowths of Metal Films on Single Crystal Inorganic Substrates

By H. R. THIRSK

Department of Physical Chemistry, King's College, University of Durham,
Newcastle-on-Tyne

MS. received 24th April 1950

ABSTRACT. Some new examples of metal crystal orientations on inorganic substrates are described. They include silver and iron on mica, platinum on the cube face of potassium chloride, iron and silver on the (111) face of potassium chloride, silver and nickel on the cube face of magnesium oxide. Some of the problems involved in the study of the types of crystal growth are discussed briefly for the cases examined.

§ 1. INTRODUCTION

EXAMPLES of the structures of thin evaporated films formed on crystalline substrates have been described by a number of workers. In a recent paper by van der Merwe (1949) concerned more particularly with the nature of the orientating forces at the metal substrate interface, a useful table was given summarizing a great amount of experimental data on epitaxial growth. In view of the value of such information, particularly as a source of data for theoretical studies, it has been thought worth while to describe in this note further examples from an electron diffraction investigation supplementing the data recorded there.

§ 2. EXPERIMENTAL

The following technique was used in the preparation of the evaporated metal films. The ends of a silica tube, 4 cm. in diameter and about 40 cm. long, were fixed by means of Apiezon wax into brass annuli with ground faces against which two brass end caps could be joined using Apiezon M grease on flat metal to metal faces. The leads for a filament, acting as a source of metal, and a thermocouple, registering the temperature of the substrate, were taken through one cap. The tube was evacuated through a wide tube in the second brass cap by means of a mercury diffusion pump backed by a Hyvac rotary oil pump. A mercury vapour trap was placed between the silica tube and the diffusion pump. The substrate on which the deposit was to be formed was heated by means of a furnace surrounding the silica tube.

Silver, nickel, platinum and iron deposits were obtained by evaporation from a filament of the pure metal. The filament was first flashed *in vacuo* and the silica tube heated over a suitable length to drive off adsorbed air. The

degassed substrate, having been brought into position near the filament, was raised to a desired temperature by localized heating, after which the filament was heated electrically until a deposit of metal could be observed forming slowly on the substrate.

The following single crystal substrates were used during the work. Magnesium oxide in the form of single crystals with freshly cleaved cubic faces about 1 cm^2 in area; (111) faces of potassium chloride prepared by grinding the required face on a small cube of the substance, polishing on filter paper with water and finally very lightly etching with water; cube faces of potassium chloride prepared by cleavage; the cleavage face of calcite and the cleavage face of muscovite mica. Electron diffraction and microscope examination of (111) surfaces of KCl indicated that, as prepared, they were extremely smooth. All the electron diffraction patterns were taken on a Finch-type instrument with an accelerating voltage of approximately 50–55 kv.

Silver on Mica

Temperatures below 200°C . gave silver deposits mainly with one-dimensional $\{111\}$ orientation. At higher temperatures the rate of deposition was extremely slow and also led to the destruction of mica surface layers. The mirrors produced at about 200°C . gave two principal diffraction patterns, one when the beam was parallel to a basal axis of the mica cleavage plane, which repeated every 60° of azimuth, and a second at the azimuth 30° to each of these repetitions. These diffraction patterns show that the silver was deposited with $\{111\}$ faces in contact with the mica cleavage face in two definite azimuth orientations, namely, with a $[110]$ lattice row parallel to either of the two basal axes of the mica, one being obtained from the other by rotation through 180° about the (111) plane normal. This corresponds to crystals in four different orientations similar to the orientations described by Rüdiger (1937) for gold and palladium and by Andrashenko, and Tyapkina and Dunkov (1948) for silver on mica.

In Figure 1 (Plate I) the beam was in a direction parallel to the $[1\bar{1}0]$ azimuth of one of the (111) Ag faces. A strong refractive index effect was present with most specimens. Indexing was carried out using the reciprocal lattice diagrams reproduced by Wilman (1940), but with the assumption that the spots on the $\frac{1}{2}$ in. diagonals were the product of double diffraction and submicroscopical twinning rather than due to the presence of a 'buffer' lattice as suggested by Menzer (1938). Closely analogous patterns to silver on mica have been obtained from nickel oxide evaporated on to the basal plane of sapphire by Thirsk and Whitmore (1940), where a complete indexed pattern applicable to the present case is given for the $[1\bar{1}0]$ azimuth.

Many of the specimens of silver on mica show a further two-dimensional orientation with a cube parallel to the main basal axis, and a cube face parallel to the mica cleavage face. The orientation was best observed on specimens with the mica substrate heated to 250° – 350°C . The original of Figure 1 shows the $[\bar{1}00]$ azimuth, (100) face diffractions.

Iron on Mica

Iron was evaporated on to mica substrate kept at temperatures between 150° and 400° , the iron filament being at bright red heat. The iron was cleaned initially by rubbing with 0000 emery and care was taken to free the iron from

oxide by preheating the filament electrically in a vacuum. No two-dimensional orientation was observed, all the patterns being due to one-dimensional orientation, $\{110\}$ and $\{111\}$, the $\{110\}$ orientation predominating. All the evaporation experiments gave excellent mirrors, which subsequently exposure showed to be remarkably resistant to rusting.

Silver on the (111) Face of Potassium Chloride

For purposes of comparison with the behaviour of silver evaporated on pseudo-hexagonal muscovite cleavage faces silver was evaporated on to the ground and etched (111) face of potassium chloride. A two-dimensional orientation is known for the (100) face.

With the substrate at 200–300° c. a pattern similar to that of silver on mica was obtained, indicating the presence of the same hexagonal orientations.

Iron on the (111) Face of Potassium Chloride

Although Brück (1936) found it comparatively easy to produce two-dimensional orientations of iron on the cleavage face of rock salt only one-dimensional orientations were observed with iron on the (111) potassium chloride face and then only with the substrate at temperatures higher than 200° c. A strong $\{110\}$ and some $\{200\}$ one-dimensional orientation was present.

Platinum on to a (100) Potassium Chloride Face

A freshly cleaved specimen of potassium chloride was maintained at a temperature of approximately 250° c., and platinum evaporated from a filament at about white heat. Several hours were required for the deposition of a film of appreciable thickness. Patterns were taken with the beam along the $[100]$ and $[1\bar{1}0]$ azimuth of the cube face. Diffractions due to the potassium chloride substrate could clearly be seen. The patterns showed that the platinum films were of a normal untwinned single crystal structure, the cube of the platinum edge being parallel to that of the cube edge of the substrate. The $[100]$ azimuth pattern is shown in Figure 2 (Plate I).

Silver on Magnesium Oxide

Magnesium oxide offered interesting properties as a substrate since, although cubic and of the sodium chloride type, it is regarded as differing from a purely ionic crystal (Wells 1947). Silver was evaporated on to the (100) cleavage face maintained at temperatures of from 400° to dull-red heat. Below 400° the deposits were mainly random. The chief orientation was $\{100\}$ with the cube edge parallel to the cube edges of the magnesium oxide. A representative pattern is shown in Figure 3 with the beam along the $\{100\}$ azimuth of the magnesium oxide. Figure 3 is a pattern due to a $\{111\}$ twinned silver lattice $[100]$ azimuth. In addition there are present diffractions corresponding to the one-dimensional orientations $\{211\}$ strong and $\{111\}$ relatively weak. The $\{211\}$ one-dimensional orientation has often been observed when face-centred cubic systems are evaporated on to comparatively inert substrates.

Nickel on Magnesium Oxide

Nickel was evaporated from a thin filament previously flashed *in vacuo* to free from oxide and then maintained at bright-red heat. No orientation of the films

was found with the magnesium oxide substrates at temperatures below 500° c. and above that temperature the orientations observed were all of one degree. Figure 4 shows that the principal one-dimensional orientation are {110} strong, {200} fairly strong, {110} fairly weak, {111} weak.

Silver on Calcite

At temperatures above 400° c. two-dimensional orientations of silver on calcite were found similar to those reported by Rüdiger (1937) for Ag, Au and Pd with {110} and {111} faces in contact with the calcite cleavage faces (Figure 13 of his paper). Below 400° c. extremely complex patterns were obtained showing however the presence of fairly well developed crystals.

The results described have been summarized in the Table.

Metal	Substrate	Comment
Silver	Muscovite mica	Below 200° c. mainly {111} one-dimensional orientation. Above 200° c. principally four {111} two-dimensional orientations [110] lattice row parallel to either of the axes in the mica basal plane. Between 250°–300° c. An additional well-marked {100} two-dimensional orientation with the cube edge of the silver parallel to the axis in the basal plane of the mica. Evidence of sub-microscopical twinning. Andrashenko <i>et al.</i> (1948) report the same cube face orientation.
Fe	Muscovite mica	Between 150° and 400° c. {110}, {111} and {110} one-dimensional orientation only. Below 150° random patterns.
Ag	(111) KCl	Between 200°–300° c. Behaviour very similar to silver on mica at the same temperatures.
Fe	(111) KCl	No preferred orientation at temperatures less than 200° c. At 350° c. a strong {110} and some {200} one-dimensional orientation only. Compare with Fe on mica and contrast with the easily obtained two-dimensional orientation observed by Brück (1936) on the cleavage face of rock salt.
Ag	(100) MgO	Below 400° c. Deposits mainly random. 400° c. to red heat. Chief orientation cubic face and cube edge parallel to the cube face and cube edge of the magnesium oxide. Evidence of twinning even at the high temperatures. Strong {211} and weak {111} one-dimensional orientations.
Ni	(100) MgO	Below 500° c. Deposits random. Above 500° c. Orientation of one-dimensional {100} strong, {200} fairly strong, {110} fairly weak, {111} weak.
Silver	Calcite cleavage face	Below 400° c. Very complex patterns probably due to a number of fairly large crystallites in slightly differing orientations. Above 400° c. Patterns confirming Rüdiger's results (1937) but also showing evidence of twinning.

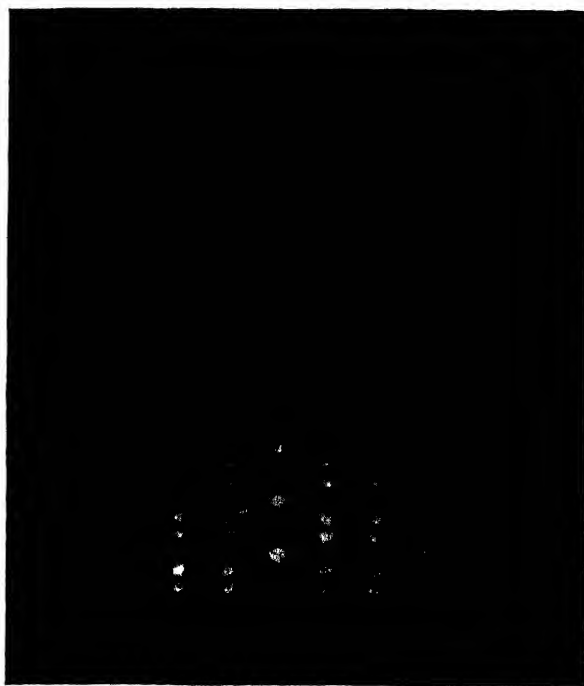


Figure 1. Silver on a mica cleavage face. Beam parallel to the $[1\bar{1}0]$ azimuth of one of the (111) Ag faces and the basal axis of the mica.

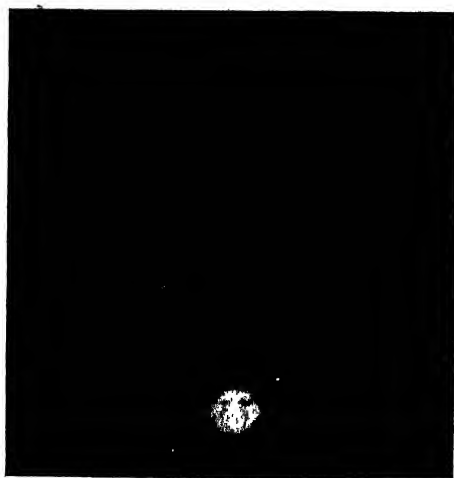


Figure 2. Platinum on a potassium chloride cleavage face. The beam parallel to the cube edge azimuth of the platinum and potassium chloride.

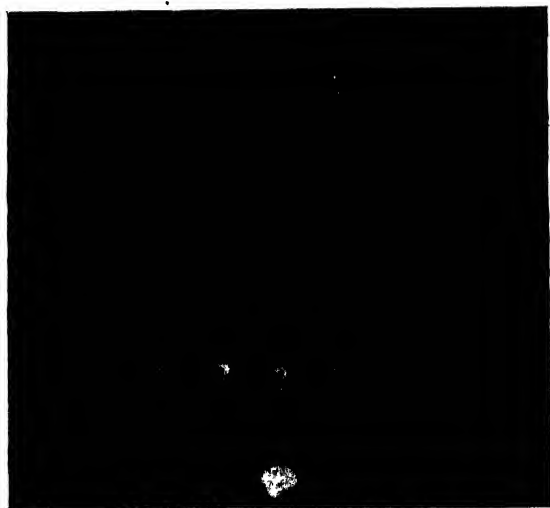


Figure 3. Silver on the cleavage face of magnesium oxide. Beam parallel to the cube edge of the principal silver crystals and the cube edge of the substrate crystal.

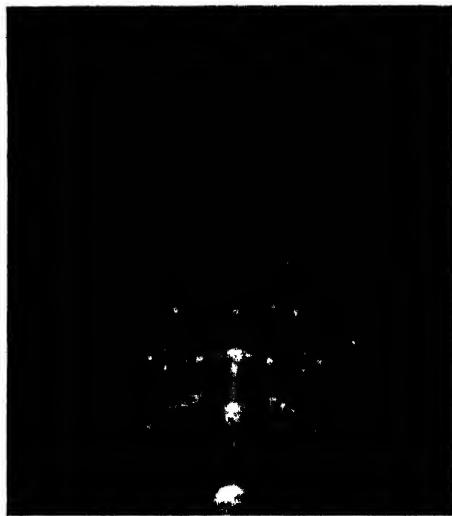


Figure 4. Nickel on the cleavage face of magnesium oxide.

§ 3. DISCUSSION

The intention of this paper is primarily to describe some additional examples of the orientations assumed by metal films evaporated on to single crystal substrates, and these have been summarized in the Table; but in examining the results, difficulties of interpretation arise, the nature of which require some discussion. We may summarize them as follows: first, the extent to which the film is free from discontinuities, not merely in the sense of local dislocations, but in the likelihood that the deposit is formed of separate crystals of small size having uniform orientation; the extent to which the interfacial forces, assumed to be controlling the nature of the orientation, may have become a true chemical bonding; and finally, the question of how far one may consider that closeness of fit of the interatomic distances of the adjacent faces of metal and substrate deduced from the bulk orientation may have been decisive in the choice of the final atomic arrangement.

Electron diffraction cannot of itself provide a complete answer to the first problem. For a completely satisfactory examination recourse must be had to a supplementary technique using the electron microscope or, in certain problems, multiple beam interferometry. Hass (1942) has shown by electron microscopy that silver films grown on the (100) face of rock salt when less than 40 Å. thick are in the form of isolated granules about 100 to 150 Å. long. The diffraction patterns from the platinum films strongly suggest a similar structure. In addition a complete absence of any diffractions due to submicroscopical twinning is noted, although the slight arcing in patterns with the beam along the cube edge and cube face diagonal suggest that the crystals are extremely thin in a direction normal to the surface plane, and usually it is in this thin layer adjacent to the substrate that the strong evidence for twinning with face-centred cubic metals is found. There is a considerable difference between the a_0 spacing for platinum and potassium chloride: $a_0(\text{Pt}) = 3.916 \text{ Å.}$, $a_0(\text{KCl}) = 6.278 \text{ Å.}$

Chemical interaction is in some cases an important factor in ensuring a surface orientation. Willems (see van der Merwe 1949 for reference) has quoted a number of examples, but a result particularly relevant to the work described here is an example studied by Thirsk and Whitmore (1940) concerning the orientation of Fe_2O_3 on a magnesium oxide substrate with the observed formation of an intermediate layer of spinel. It appears from a consideration of the behaviour of nickel on magnesium oxide and iron on mica, that chemical interaction may under certain conditions prevent rather than assist the production of a two-dimensional orientation. In these two cases chemical bonding may prevent further movement after condensation of the nickel and iron on the substrate. With analogous experiments with evaporated silver, the surface mobility is retained in the absence of chemical bonding.

The behaviour of magnesium oxide is particularly interesting owing to the diversity of its behaviour as an active substrate. It has been used as such by Vineyard (1942) for the ionic crystals LiF, NaF and NaCl deposited from solution and from the melt. There was an easily produced two-dimensional orientation only in the case of LiF, where the lattice fit is very close. In the present work the lattice fit of Ag on magnesium oxide is good ($a_0(\text{Ag}) = 4.077 \text{ Å.}$, $a_0(\text{MgO}) = 4.203 \text{ Å.}$). We can summarize the behaviour of magnesium oxide as an active substrate. From Vineyard's work a controlling factor for ionic deposits is apparently a

close lattice fit. From previous work by the author with a non-ionic deposit Fe_2O_3 , chemical interaction and the formation of an intermediate compound can produce a two-dimensional orientation. The spinel interfacial compound shows signs of twinning. With nickel deposits reported in this paper chemical bonding may occur, two-dimensional orientation being prevented by the limitations thus imposed on the mobility of the nickel atoms in the first deposited layer. With silver a moderate degree of two-dimensional orientation is obtained at high temperatures due to Van der Waals surface forces but presumably without chemical bonding. There is evidence again of submicroscopical twinning.

Wilman (1940) has shown that silver films separated from the substrates and showing twinning spots may readily be annealed by heating to 500°C . Temperatures higher than this failed to produce annealing with the films of silver on magnesium oxide, suggesting that the surface forces at these higher temperatures must still be very considerable.

The patterns from silver on mica and on the (111) potassium chloride face are very similar to those described by Rüdiger (1937), but the extra diffractions are interpreted as being due to submicroscopical twinning and double diffraction.

Additional details of the lattice fits at the interface of the deposit and substrate implied by the orientation of the bulk deposit are as follows: for silver ($a_0 = 4.06 \text{ \AA}$) having a {111} face orientation on muscovite mica a percentage misfit of 10% must exist, using van der Merwe's nomenclature, if we assume that the silver ions fit into the hexagonal arrangement of spaces in the oxygen network as suggested by Andrashenko and co-workers (1948), or -3% if the position of the OH' groups in the mica with a hexagonal spacing of 3.0 \AA are of importance in deciding the orientation (for the structure of mica, see Jackson and West 1930). With silver on the (111) potassium chloride face ($a_0 = 6.06 \text{ \AA}$) the misfit is of the order of 30%. Rhodin (1949) examining the orientation of aluminium on the (111) face of LiF and NaCl associates the poorer fit on the NaCl face with a greater disparity in the principal lattice dimensions. It was pointed out by Thomson (1948) in summarizing data on the orientations assumed by metal films that they are frequently not those giving the apparently most favourable lattice. Van der Merwe suggests that Menzer's geometrical picture (see Menzer 1936, 1938) of the establishment of a final orientation through the growing together of intermediate layers of different orientation having a good lattice fit at the substrate surface is a plausible treatment of the problem. This hypothesis explains the existence of certain extra spots in the diffraction patterns, but it has been shown that they may also be explained by less elaborate postulates concerning the nature of the deposit (Wilman 1940, Thirsk and Whitmore 1940). An alternative process would be through the primary formation of an approximation to a hexagonal close-packed layer followed by a change through adjacent layers to the final orientation, any strain produced being relieved by slip processes, for instance those giving rise with face-centred cubic metals to submicroscopical twins. With a process of this kind the nature of the symmetry of the rigid substrate face by defining positions of general lattice fit, having a symmetry similar to that of the substrate face, could cause the appearance of a similar symmetry in its final film. This is then maintained even though the agreement in spacing atom for atom along the main lattice rows at the interface is poor and liable to give rise to a large number of local dislocations in the films deposited.

ACKNOWLEDGMENTS

The author would like to record his thanks to Professor G. I. Finch for his interest and Dr. H. Wilman and Dr. E. J. Whitmore for much helpful discussion during the period when the experimental work described in this paper was being carried out.

REFERENCES

- ANDRASHENKO, H. V., TYAPKINA, V. V., and DUNKOV, P. D., 1948, *Doklady Acad. Nauk, S.S.S.R.*, **59**, 1113.
BRÜCK, L., 1936, *Ann. Phys., Lpz.*, **26**, 233.
GOCHE, O., and WILMAN, H., 1939, *Proc. Phys. Soc.*, **51**, 625.
HASS, G., 1942, *Kolloid Z.*, **100**, 230.
JACKSON, W. W., and WEST, I., 1930, *Z. Kristallogr.*, **76**, 211.
VAN DER MERWE, J. H., 1949, *Discussions on Crystal Growth* (London: Faraday Society).
MENZER, G., 1936, *Naturwissenschaften*, **26**, 385; 1938, *Z. Kristallogr.*, **99**, 378, 410.
RHODIN, T. N., 1949, *Discussions on Crystal Growth* (London: Faraday Society), p. 215.
RÜDIGER, O., 1937, *Ann. Phys., Lpz.*, **30**, 505.
SHIRAI, S., 1937, *Proc. Phys.-Math. Soc., Japan*, **19**, 937; 1938, *Ibid.*, **20**, 855; 1939, *Ibid.*, **21**, 800; 1941, *Ibid.*, **23**, 12.
THIRSK, H. R., and WHITMORE, E. J., 1940, *Trans. Faraday Soc.*, **36**, 565, 862.
THOMSON, G. P., 1948, *Proc. Phys. Soc. A*, **61**, 403.
VINEYARD, S. M., 1942, *Phys. Rev.*, **61**, 100.
WELLS, A. F., 1947, *Structural Inorganic Chemistry* (Oxford: University Press).
WILMAN, H., 1940, *Proc. Phys. Soc.*, **52**, 323.

Surface Effects in Creep of Cadmium Crystals

By D. J. PHILLIPS AND N. THOMPSON

H. H. Wills Physical Laboratory, University of Bristol

Communicated by N. F. Mott; MS. received 31st May 1950

ABSTRACT. A recording rate-of-strain meter has been used to investigate the changes in creep rate when the chemical environment of a stressed cadmium crystal is altered. Experiments were made with aqueous solutions of a number of inorganic salts of cadmium. The observations are explicable in terms of variations in the thickness of a surface film of cadmium hydroxide, the presence of such a film having the effect of reducing the creep rate by an amount depending on its thickness. Measurements were also made on the effect of very thin films, formed by immersion in distilled water; these were estimated to be about 10^{-8} to 10^{-7} cm. thick. When the films were removed with dilute sulphuric acid a sudden small strain increment was observed, the magnitude of which depended on the film thickness. Possible explanations are discussed.

§1. INTRODUCTION

THE work to be described was suggested by certain effects reported by Andrade and Randall (1948). In their work, clean single crystals of cadmium were made to extend uniformly at a rate of about 1.7×10^{-4} per minute, and were then surrounded by various inorganic salt solutions (mainly salts of cadmium). They found that the rate of strain suffered changes which were in some cases very large. For example, the application of cadmium sulphate solution increased the strain rate by a factor of 4; cadmium nitrate gave at first an increased rate, followed by a complete stoppage; with zinc nitrate, immediate hardening occurred.

Effects of surface conditions upon the strength of metal crystals have also been reported by Roscoe (1936), by Cottrell and Gibbons (1948) and by Harper and Cottrell (1950) who were concerned with oxide films. A considerable amount of work has also been done in Russia by the Rehbinder school. The fracture and cleavage strengths of mica and calcite, and the resistance to plastic deformation of metals (polycrystals and single crystals) were all found to be reduced when various surface-active substances (e.g. higher alcohols) were added to the medium surrounding the specimen (Rehbinder and co-workers 1931–1947). Some doubt however, exists as to the validity of some of this work (Kemsley 1949) and the question has been discussed by Harper and Cottrell (1950).

In order to study the variation in strength (i.e. resistance to plastic flow) of a crystal, two methods suggest themselves:

(a) The measurement of critical shear stress. A number of uniform crystals, of the same orientation, would be given various surface treatments and the stress required to produce a known, small strain rate measured.

(b) A given specimen could be made to extend under constant load, and the various surfaces prepared while the test was in progress. The effects of these changes on the rate of strain could then be studied.

The first method gives quantitative data: even with the best techniques, however, the measurement of changes of critical stress of 1% or less would require statistical experiments upon numerous samples. Moreover, the method is obviously not suited to continuous measurements. In some of the experiments to be described, surface conditions were not stable and the use of this method would have been impossible. The second method was chosen for the present work. It has the disadvantages that the crystal is subjected to strain-hardening during the experiment, and that it is difficult to obtain a quantitative measure of the effects. For the *detection* of the effects, however, the method is very sensitive.

In practical convenience the second method is much to be preferred. In one experiment many surface changes can be investigated (as many as 90 separate tests have been made). Further, since work-hardening must in any case be present, there is no reason why the same crystal should not be used for many experiments. In the present work, a crystal was not discarded until the limit of plastic flow was approached (about 40% extension). By its nature, the method is not affected by some factors which would be important in the first method. Changes in temperature or stress, for example, provided that they occur slowly, do not affect the results since only rapid changes in strain rate were measured. Nor was it necessary to take any great precautions regarding the initial alignment and straightness of the specimen, since fluctuations in creep rate due to these factors disappear after a small extension.

§ 2. EXPERIMENTAL DETAILS

The cadmium used was kindly supplied by the National Smelting Co. The impurities were as follows: lead 0.007%, zinc less than 0.01%, thallium 0.0005%, tin less than 0.0005%, copper less than 0.0001%, calcium trace: total less than 0.018%.

Crystals were grown in air by the method of Andrade and Roscoe (1937). They were of slightly oval cross section, of mean diameter 1.9 mm., and 60 cm.

long. They were cut into 10 cm. lengths for the experiments. The crystals, as grown, were somewhat oxidized. They were cleaned by electropolishing for ten seconds in 50% orthophosphoric acid, at a current density of about 1.5 amp/cm^2 .

The results quoted in this paper were all obtained with portions of one crystal, in which the angle between the slip plane and the specimen axis was 25° , and the angle between the slip direction and the specimen axis was 36° . The crystal was supported vertically and loaded by hooks soldered to its ends, the loads being applied by means of a weight and lever. The strain was measured by recording photographically the movement of a spot of light reflected from a mirror fixed to the lever. The magnification was 16 times.

Since the work primarily required observation of the rate of strain, a device to record this directly was also used. A thread affixed to the lever passed over the pivot-shaft of a moving-coil milliammeter, and ended in a small weight. Movement of the lever caused the coil to rotate, and the resulting induced current actuated a galvanometer. The galvanometer deflection was therefore proportional to the velocity of the lever, and hence to the rate of strain of the specimen. This is true, of course, only if the strain rate is approximately constant. If the specimen elongates discontinuously in a time short compared with the galvanometer period, the system effectively integrates, and a transient deflection is obtained, proportional to the strain increment. A photographic record of the strain rate was made on the same camera as was used to record the strain. The velocity sensitivity of the arrangement used was 800 cm. deflection for 1 cm/sec. specimen velocity, the displacement sensitivity for sudden extensions was 6,300 cm. per cm. extension of the specimen.

The crystals were subjected to the action of the various reagents by directing a fine jet of solution on to the upper end of the specimen. Two jets were used, either of which could be directed upon the specimen at will by a slight rotation of their support. The rate of flow was $0.5 \text{ cm}^3/\text{sec}$. In the work on thin films (§4), where it was thought that the forces exerted by the jets on the crystal might vitiate the results, the jets impinged upon a small aluminium plate, and were conducted to the specimen by a projection of this plate which almost touched it. The liquids used were always within 0.5°C . of room temperature, and careful tests showed that errors due to thermal changes were not important.

It was found that any change in the composition of the solution surrounding the metal usually produced a change in the creep rate. The extent, and even the sign, of the change in rate depended not only on the nature of the new solution, but also on the previous history of the specimen. It appears that the changes in creep rate are to be correlated with changes in the thickness and/or composition of the film of hydroxide* which forms on the metal surface under these conditions. The experimental results fall conveniently into two groups distinguished by the order of magnitude of the thickness of the surface films involved. The results are presented separately in §§3 and 4.

* No attempts were made to determine the exact composition of the surface films formed. When they were thick enough to be visible they were white, and turned into the characteristic brown oxide on warming. Mellor (1929) speaks of films of a 'hydrated oxide' being formed on cadmium in water: for convenience the films are referred to as 'hydroxide' throughout. (Since the above was written Mentor and Hall (1950) have reported the occurrence of a crystalline overgrowth on oxide-coated cadmium crystals immersed in cadmium chloride or nitrate: from electron diffraction photographs this was identified as cadmium hydroxide.)

§3. RESULTS: THICK HYDROXIDE FILMS

To demonstrate the effects of relatively thick films of cadmium hydroxide, an electrolytic method of preparation was used. The specimens were anodically oxidized in an electrolyte of potassium hydroxide at a current density of about 1.4 amp/cm^2 . The hydroxide film builds up in a few seconds to an equilibrium thickness of about 10^{-4} cm . (The film thicknesses quoted throughout were estimated by weighing, the density of the film being taken to be the same as that of bulk cadmium hydroxide. Because of this assumption the values of absolute thickness are liable to be in error, perhaps by a factor of two. The relative accuracy, however, is to within a few per cent.)

Figure 1, which is copied from one of the original photographic records, shows the effect upon the creep rate when such a film is removed. Before the point B the hydroxide-covered crystal was extending uniformly in distilled water at a low rate (about 3×10^{-5} per min.). At B, 2% sulphuric acid was directed upon the crystal. Cadmium metal is not rapidly attacked by acid of this concentration: a separate experiment showed that the rate of solution of the metal was about $4.4 \times 10^{-6} \text{ cm. depth per minute}$. The hydroxide film, on the other hand, was removed within one second by conversion to the soluble sulphate. The rate of extension immediately increased by a factor of about 60; subsequently it slowly decreased again, due to the hardening of the crystal. At the point C, distilled water was substituted for the acid, with no sensible effect upon the strain rate. With more dilute acid (0.1%) the removal of the film and the consequent rise in strain rate took place more slowly, rising from 10^{-4} per minute to a maximum of 1.5×10^{-3} per min. in 1.3 minutes.

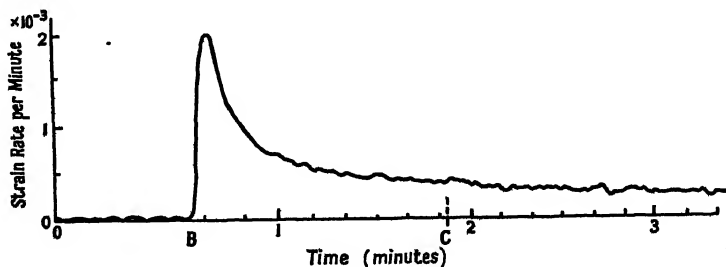


Figure 1.

Immersion of an electropolished crystal in a solution of cadmium nitrate was found to deposit hydroxide films and to reduce the creep rate. The removal of such films with sulphuric acid while creep was in progress produced effects entirely similar to those mentioned above. The films were easily visible; immersion for four minutes in a semi-normal solution produced a film $4 \times 10^{-5} \text{ cm. thick}$. The effect of depositing a film upon a clean crystal while creep was in progress is shown in Figure 2. The crystal was extending in distilled water. At the point A semi-normal cadmium nitrate was applied, and the strain rate immediately began to decrease. At B 2% sulphuric acid was substituted for the nitrate. The removal of the film built up during the application of the cadmium nitrate solution caused the large increase in strain rate shown.

On the other hand if the crystal were already covered with a 'thick' hydroxide film, prepared by anodic oxidation as previously described, application of the same reagent (cadmium nitrate) caused an increase in the creep rate. This behaviour is

shown in Figure 3. Curve *b* shows the effects upon the rate of strain when a semi-normal solution of cadmium nitrate was applied (at $t=0$) the crystal having been previously coated electrolytically with hydroxide. The strain rate rose by a factor of 6.3 in 2 minutes. Subsidiary weighing experiments showed that under these conditions the thickness of the film was reduced. At $t=147$ seconds, 2%

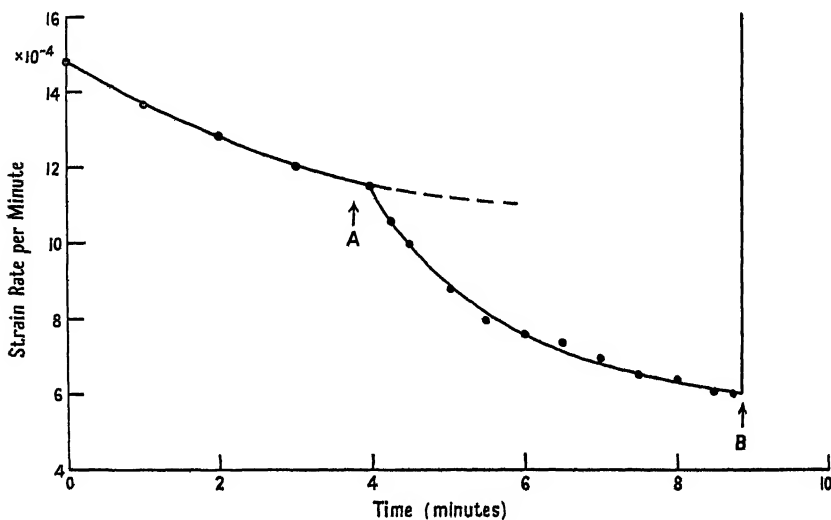


Figure 2.

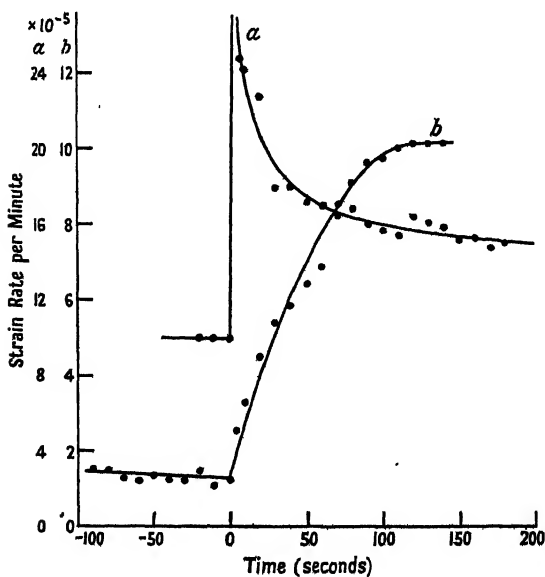


Figure 3.

sulphuric acid was applied. The effect is shown in curve *a*. This has been drawn so that the instant of application again falls at the time origin. The rapid removal of the remainder of the hydroxide film causes a typical rapid increase in the strain rate followed by hardening.

Various other substances were found to behave in a similar manner, with partial removal of a pre-existing thick hydroxide film; the salts investigated were a commercial cadmium plating solution (Canning 'Zonax') and semi-normal solutions of cadmium chloride, nitrate and sulphate. The speed with which the film was removed varied in this order (from about 3 seconds for the first to 100 seconds for the last); the records obtained showed an increased creep rate when the solution was applied. The magnitude of the increase was less for slow solution of the film than for rapid solution, probably because of the work hardening taking place at the same time. The plating solution gave a curve similar to that obtained with 2% sulphuric acid while cadmium sulphate gave a curve similar to that with 0.1 % acid. In the case of the chloride solution it was verified that it caused the deposition of a film on an electropolished crystal, and reduced the creep rate.

All these observations are consistent with the view that, other things remaining constant, the creep rate is reduced by the presence of a film of hydroxide on the crystal, that the effect is more marked the thicker the film, that there is an 'equilibrium' thickness of film which depends on the composition of the solution in contact with it, that the substitution of one solution for another will increase or decrease the creep rate according as the film thickness decreases or increases, and that the rate of the change depends on the concentration of the solutions used.

§ 4. RESULTS: THIN FILMS

The records of strain rate show irregular fluctuations (see Figure 1) which may be due to friction in the measuring system, or may possibly represent real variations. In any case their presence means that small changes in strain rate cannot be detected with certainty. However, it was observed that when a thin film is removed rapidly from the surface of the metal, although the resulting change in creep rate may be too small to be detected, there is in addition a sudden extension at the instant of removal. This is readily observed and, owing to the integrating property of the strain-rate recorder already mentioned, the resulting transient deflection of the galvanometer is in this case proportional to the strain increment. The removal of such a film appears to be very similar in its effects upon the creep rate to a sudden small increase in load. It was found that films whose thickness was estimated to be only 20 Å. could be detected in this way.

The films investigated were produced by immersing the crystal in distilled water. Cadmium immersed in water with access of air is known to form a hydrated oxide (Mellor 1929). By weighing, it was estimated that the film deposited in ten minutes had a thickness of about 3×10^{-6} cm. Such films were rapidly removed by sulphuric acid. The procedure was to allow the crystal to creep under constant load for a few minutes until the creep rate had ceased to fall rapidly. It was then exposed alternately to distilled water and to 2% sulphuric acid. A transient deflection of the galvanometer occurred at each change from water to acid, but none at the opposite interchange, thus suggesting that a film had been formed during the exposure of the crystal to water. This is shown in Figure 4, in which the changes from acid to water are marked by single ordinates, and those from water to acid by double ordinates.

If the changes of liquid were repeated many times at regular intervals—say once per minute—the magnitude of the transient showed a progressive fall as the crystal hardened. In order to allow for this, the applied load was momentarily increased at frequent intervals during the experiment. These load increments were,

typically, 1% of the steady load and were applied for 3 seconds. They gave rise to a sudden strain increment, which appeared on the strain-rate record as a transient similar to that produced by removing the thin hydroxide film. By comparing the heights of the transients due to acid application with those due to a standard load increment, the effects of strain-hardening could be allowed for.

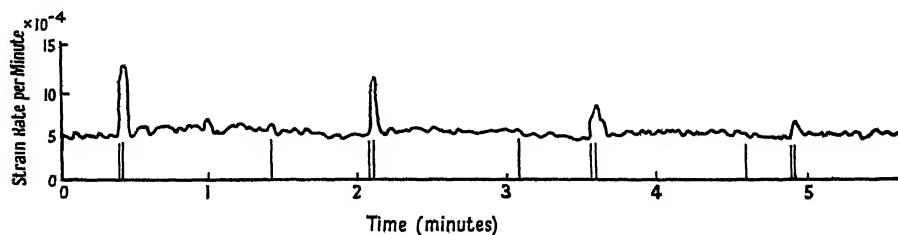


Figure 4.

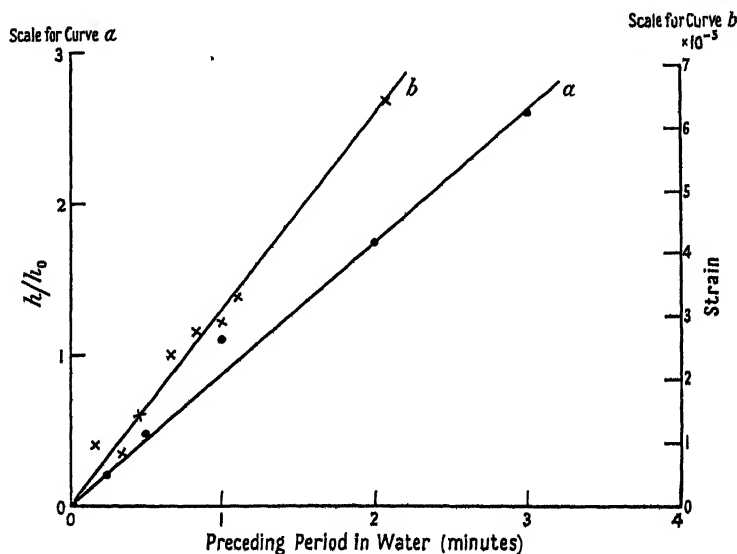


Figure 5.

In this way it was found that the magnitude of the sudden extension due to acid application increases when the duration of the previous immersion in water is increased. This dependence is shown in Figure 5 curve *a* in which ordinates are the ratio of the corresponding transient galvanometer deflection to the 'standard' deflection produced by a load increment of 0.2%. Figure 5 curve *b* shows a similar set of observations, obtained from a specimen which showed only very little strain-hardening during the time occupied by the measurements: the absolute magnitude of the transients can accordingly be given without ambiguity. The simplest interpretation of Figure 5 is that the film thickness increases linearly with time at least for the times shown. If this is true, then using the result already quoted for the estimated thickness of the film after 10 minutes immersion, we may conclude that the removal of a film about 10^{-7} cm. thick produced a measurable effect. This consisted of a sudden increase in strain of order of magnitude 10^{-5} ; there may possibly have been also an increase in strain rate, but if so it was too small to be detected.

§ 5. DISCUSSION

In view of the results reported in §§ 3 and 4 it appears probable that the results of Andrade and Randall mentioned in the introduction would be explicable in terms of surface films formed or removed by the inorganic reagents which they used. It is also becoming clear that the conflicting results concerning the effects of organic liquids may be reconciled when due importance is attached to such films. Andrade and Randall (1949) have shown that the 'Rehbinder effect'—increase in creep rate caused by immersion in a dilute solution of oleic acid in paraffin—is only found with cadmium when the metal is originally covered with an oxide film. Similar results have been reported in more detail by Harper and Cottrell (1950). Both papers report a considerable time interval between the application of the liquid and the increased creep rate—as much as 45 minutes in some circumstances—so that the mechanism of 'removal' of the film is possibly not the same as when the hydroxide is dissolved in strong acid.

In any event there remains the problem of exactly how the presence of a thin film of oxide or hydroxide on the surface can reduce the rate of creep to such a marked extent. One cannot reject out of hand the naïve hypothesis that the film simply relieves the metal of part of the tensile load. The results already given in §§ 3 and 4 enable us to calculate the load which the film must have been supporting, and the estimates of thickness enable us to convert this to stress. The results obtained are of the order of 10^9 dynes/cm² which, although not entirely impossible, is sufficiently high to make the explanation somewhat improbable. It is interesting to note that Roscoe (1936), dealing with the effects of oxide films on cadmium, deduced a value for the strength of the film of the same order of magnitude although by a very different line of approach.

In the same way it appears unlikely that the explanation is to be found in thermochemical effects. Certainly the amounts of heat produced by the chemical reactions involved are too small to have any effect if distributed through the whole volume of the metal. But there is also the possibility that a local heating of the surface might suffice to 'trigger' a slip process which would then extend right across the crystal. Such for example would appear to be a possible explanation of the effect reported by Andrade (1945) that bombardment of the surface of a cadmium crystal with α -particles can cause a considerable increase in the creep rate. In the present experiments, however, approximate calculations show that the temperature changes involved are not sufficient to give rise to any such process.

A more promising suggestion is concerned with the influence of the surface film on sub-microscopic regions of stress-concentration near the surface of the metal. If we adopt current views, we are to describe processes of plastic deformation of metal crystals largely in terms of the movement of 'dislocations'. It can be shown that the formation of new dislocations is an improbable process for ordinary values of applied stress, and is only likely to occur where local conditions provide a stress concentration. One place where such conditions might arise is on the free surface of a single crystal, and in so far as we must consider the formation of new dislocations at all, it is not unreasonable to suppose that the process might be considerably hindered by the presence of a solid film. Similarly, dislocations already existing within the crystal and approaching a free surface at low speeds might be held up by the presence of a film on the surface: the removal of the film would release all such dislocations at once and might account for the small sudden extensions ('transients') discussed in § 4. This suggestion receives some support

from the fact that all the effects discussed in this paper are hardly noticeable with polycrystalline material, in which the importance of the free surface of the metal is eclipsed by the large number of grain boundaries which can also act as sites of possible stress concentration.

REFERENCES

- ANDRADE, E. N. DA C., 1945, *Nature, Lond.*, **156**, 113.
 ANDRADE, E. N. DA C., and RANDALL, R. F. Y., 1948, *Nature, Lond.*, **163**, 890; 1949, *Ibid.*, **164**, 1127.
 ANDRADE, E. N. DA C., and ROSCOE, R., 1937, *Proc. Phys. Soc.*, **49**, 152.
 COTTRELL, A. H., and GIBBONS, D. F., 1948, *Nature, Lond.*, **162**, 488.
 HARPER, S., and COTTRELL, A. H., 1950, *Proc. Phys. Soc. B*, **62**, 331.
 KEMSLEY, D. S., 1949, *Nature, Lond.*, **164**, 1127.
 MELLOR, J. W., 1929, *A Comprehensive Treatise on Inorganic and Theoretical Chemistry* (London: Longmans), **4**, 474.
 MENTER, J. W., and HALL, E. O., 1950, *Nature, Lond.*, **165**, 611.
 REHBINDER, P., 1931, *Z. Phys.*, **72**, 191; 1947, *Nature, Lond.*, **159**, 866.
 REHBINDER, P., KALINOWSKAYA, N., and WENSTRÖM, E., 1933, *Phys. Z. Sowjet*, **4**, 365.
 REHBINDER, P., and LICHTMAN, V. I., 1941, *C. R. Acad. Sci. U.R.S.S.*, **32**, 130.
 REHBINDER, P., LICHTMAN, V. I., and MASLENIKOV, V. M., 1941, *C. R. Acad. Sci. U.R.S.S.*, **32**, 125.
 REHBINDER, P., and LOGGINOW, G., 1941, *C. R. Acad. Sci. U.R.S.S.*, **30**, 491.
 REHBINDER, P., and WENSTRÖM, E., 1937, *Bull. Acad. Sci. U.R.S.S.*, **4**, 531; 1944, *Acta Physicochim.*, **19**, 36.
 ROSCOE, R., 1936, *Phil. Mag.*, ser. 7, **21**, 399.

Creep in a Precipitation-Hardened Alloy

BY M. DAVIS AND N. THOMPSON

H. H. Wills Physical Laboratory, University of Bristol

Communicated by N. F. Mott; MS. received 15th March 1950

ABSTRACT. Measurements have been made of the creep shown by polycrystalline wires of a hardened alloy (Cu+3% Ag) both at room temperature and 90° K. The results are interpreted in terms of the Mott-Nabarro theory of transient creep. The agreement between theory and experiment is qualitatively good, and the quantitative discrepancies suggest directions in which the existing theory might be refined.

§ 1. INTRODUCTION

MOTT and Nabarro (1948) have published a theory of transient creep in metals based on the idea of the gradual exhaustion of a supply of dislocations already present in the metal crystal at the beginning of the test. The quantitative conclusions of the theory are, for reasons more fully explained in § 2, applicable in their simple form to a 'hard' alloy under small stresses. Since no data were available in a convenient form for comparison with the theory, the experiments described in this paper were undertaken to provide such data; the results obtained have accordingly been interpreted throughout in terms of the concepts of this theory.

For convenience we give in §2 a brief resumé of the theory. The mode of presentation is not that of the original paper but seems better adapted to the interpretation of the subsequent experimental results.* A brief account of some of the experimental details is given in §3 and the results obtained are presented in §4.

§ 2. THEORY

A fundamental assumption is that any real metal crystal contains a large number of dislocations present in the lattice and that the movement of these, under the action of applied stresses, constitutes the plastic deformation. Following Taylor (1934) it is further assumed that the motion of any dislocation, or loop of a dislocation, is impeded by a complex system of internal stresses arising from impurity atoms, precipitates, other dislocations, grain and mosaic boundaries, etc. Considering any one dislocation loop, one supposes that it will move only when the force on it, due to the externally applied stress, becomes greater than that due to the system of internal stresses, i.e. each loop will have a characteristic stress σ which is just sufficient to set it in motion. Consequently the system of internal stresses may be characterized by a distribution function, $N(\sigma)$ such that $N(\sigma)d\sigma$ specifies the number of dislocations per unit volume which can only just be set in motion by an applied stress between σ and $(\sigma + d\sigma)$. The variation of the function $N(\sigma)$ with σ is, of course, not known, but may well be of the form sketched in Figure 1.

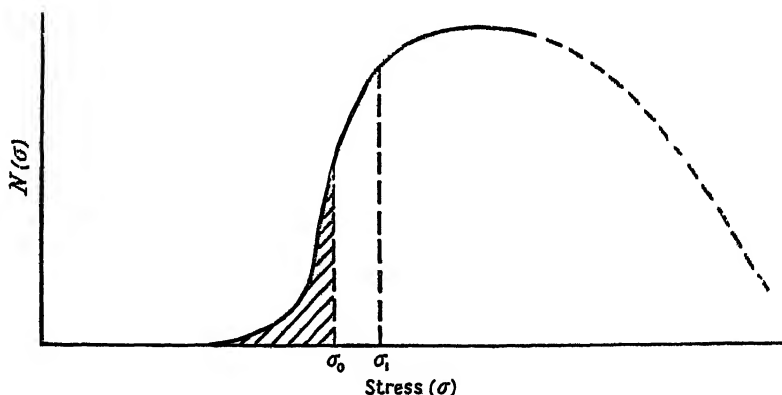


Figure 1. $N(\sigma) d\sigma$ is the number of dislocation loops per unit volume which would just move, at absolute zero, under an applied stress σ .

The application of a certain stress σ_0 will cause the immediate movement of all dislocations of characteristic stresses $\sigma \leq \sigma_0$ (i.e. those in the shaded portion of Figure 1). The immediate movement of these dislocations constitutes the 'instantaneous' plastic deformation. Dislocations which require an applied stress somewhat larger than σ_0 may possibly move in the course of time with the aid of thermal energy fluctuations. The plastic deformation so caused will constitute transient creep. Making certain plausible assumptions, Mott and Nabarro obtain the following expression for the activation energy for the motion of a loop of a dislocation, of length $\frac{1}{2}\Lambda$:

$$U(\sigma) = 0.15\sigma a\Lambda^2(1 - \sigma_0/\sigma)^{3/2}, \quad \dots\dots(1)$$

* The method of treatment is due essentially to Professor Cottrell and was presented by him to a colloquium in Birmingham in 1948; it is published here for the first time with his kind permission.

where a is the lattice parameter, σ_0 is the applied stress, and σ is the applied stress under which the loop would just move immediately.

Dislocations will not be stationary in the lattice, but will vibrate about their equilibrium positions with a frequency ν , which is estimated theoretically to be of the order of 10^8 or 10^9 per second. (Much of the discussion of the results of this investigation is concerned with the value of this quantity.) Thus the chance α dt that the dislocation loop moves forward in a time dt is given by

$$\alpha = \nu \exp \{ - U(\sigma) / kT \}. \quad \dots\dots(2)$$

Eliminating $U(\sigma)$ between (1) and (2), we have

$$\alpha = \nu \exp \left\{ - \frac{B}{kT} \left(1 - \frac{\sigma_0}{\sigma} \right)^{3/2} \right\}, \quad \dots\dots(3)$$

where B is written for $0.15\sigma a\Lambda^2$.

It follows from the discussion in the original paper that the chance of a dislocation moving backwards (i.e. against the applied stress) is negligibly small. It is further assumed that once a dislocation has moved forward in this manner, the value of U appropriate to its new position will usually be too large for further motion to be possible, so that it has effectively been used up. Thus the number of dislocations in a stress range σ to $(\sigma + d\sigma)$ will decay according to the law

$$N(\sigma, t) = N(\sigma, 0) \exp(-\alpha t).$$

At any instant t , therefore, the distribution function for the dislocations remaining can be calculated from the original distribution function by multiplying it by $\exp(-\alpha t)$, where α depends on σ according to equation (3). Insertion of reasonable numerical values shows that the function $\exp(-\alpha t)$ changes from a very small value to a value of almost unity over a short range of σ . The point of inflection on this step-like function is closely given by $\alpha t = 1$. As time proceeds, the position of this point of inflection changes in the direction of increasing σ . Without introducing serious errors, the function may be replaced by a sudden step, which advances along the axis of σ . The rate of advance is given by setting $\alpha = 1/t$ in equation (3). If σ_1 denotes the stress to which the step has advanced in a time t , we get, on rearrangement

$$\left\{ \frac{kT \ln \nu t_1}{B} \right\}^{2/3} = \frac{\sigma_1 - \sigma_0}{\sigma_1} \simeq \frac{\sigma_1 - \sigma_0}{\sigma_0} \quad \dots\dots(4)$$

if σ_1 is not very different from σ_0 .

The creep strain will depend on the number of dislocations which have moved, and if we assume that the motion of each one gives rise to an average strain v/V where V is the volume of the crystal, then the total creep strain is

$$S = v \int_{\sigma_0}^{\sigma_1} N(\sigma, 0) d\sigma.$$

If we further assume that, over the limited range of σ concerned, we may take the original distribution function $N(\sigma, 0)$ to be independent of σ and equal to N_0 , then

$$\begin{aligned} S &= v N_0 (\sigma_1 - \sigma_0) \\ &= v N_0 \sigma_0 (kT \ln \nu t / B)^{2/3}. \end{aligned} \quad \dots\dots(5)$$

Equation (5) is the same as that derived by Mott and Nabarro for the same conditions. With rather greater generality the distribution function may be taken as

$$N = N_0(1 - \sigma_0/\sigma)^m; \quad \dots\dots(6)$$

the resulting creep equation is then essentially the same as (5) except that the index $2/3$ is replaced by $2(m+1)/3$.

One feature which is essential to the validity of this argument is that the distribution $N(\sigma)$ is unchanged throughout the experiment, except by the exhaustion process under consideration. The theory will not therefore be directly applicable to any material which shows appreciable re-softening or self-annealing at the temperature of the creep test. This follows since any such process can be thought of as an increase in the number of dislocations which will move under the lower stresses.

Similarly we must also exclude, for the present, materials in which an important part of the system of internal stresses is due to the presence of the dislocations themselves; in this case the movement of dislocations in the lower ranges of stress will considerably alter the forces acting on those remaining, and the distribution function will change continually as creep proceeds. This difficulty can be minimized if we set up a more important system of internal stresses, which outweigh the effects of the varying stresses due to work hardening and are themselves unaltered by the slip process, at least for small strains. It was thought that this requirement could best be met by an alloy hardened by internal oxidation or by incipient precipitation. In addition, the observations should, for preference, be confined to comparatively small strains.

For these reasons it was decided to carry out careful creep measurements on a precipitation hardened alloy. The system copper-silver was chosen as being suitable, since the precipitation process has been fairly well investigated and is not too complex. It is suppressed completely at room temperature, so that observations can be made on a stable material. Moreover, the comparatively high melting point suggests that appreciable self-annealing is unlikely to take place at room temperature.

§3. EXPERIMENTAL

All the results given in this paper refer to polycrystalline wires of a copper-silver alloy containing 2.75 per cent by weight of silver. The copper-silver alloy was prepared from high purity, spectroscopically standardized materials and was supplied by Messrs. Johnson, Matthey in the form of hard drawn wire, approximately 1.5 mm. in diameter. Specimens 12 cm. in length were annealed for three days at the eutectic temperature (780°C.) to ensure complete solution of the silver constituent. Both this solution and subsequent ageing heat treatments were carried out in an atmosphere of hydrogen in a tubular electric furnace. Each operation was terminated by water quenching. After surveying available x-ray data, a standard ageing heat treatment of 30 minutes at 500°C. was chosen to ensure that continuous precipitation with no crystal fragmentation would occur. The polycrystalline specimens so prepared had a mean grain size of 0.1 mm.

Attempts to grow single crystals of the alloy by the Bridgman technique were unsuccessful; the crystals were rarely single and there was always a pronounced segregation of silver to one end. Some degree of success was achieved by growing copper single crystals, electroplating them with appropriate weights of silver, and

then subjecting them to a prolonged anneal at a high temperature. Alloy single crystals produced in this manner always had a slightly non-uniform silver concentration along their radius and were accordingly not considered suitable for quantitative creep measurements.* Some observations were made on them, however, and these reproduced all the characteristic features of the behaviour of the polycrystalline wires upon which most of the measurements were made. This provided confirmation for the belief that the presence of the grain boundaries would have a negligible effect on the general behaviour of the material at the temperatures concerned, which were all low compared with the melting point.

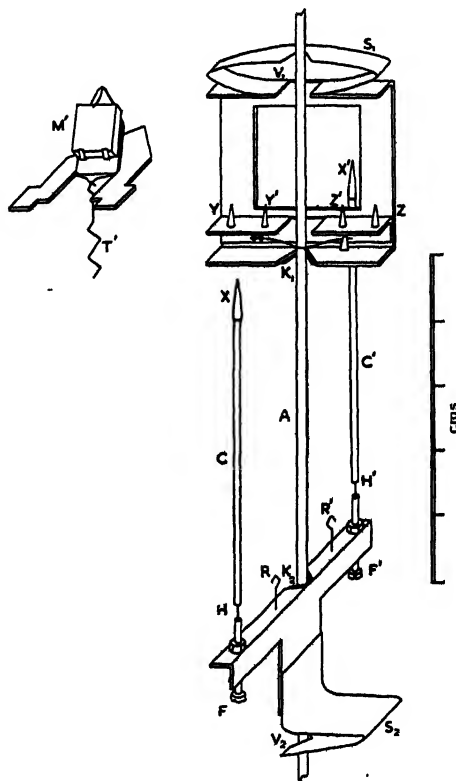


Figure 2. Extensometer; both mirrors removed and one shown inset.

The specimens were extended in a simple tensile machine, designed to apply loads up to 20 kg. As investigations were to be confined to small strains ($<1\%$) no provision was made for continuous compensation of load to maintain constancy of stress. The lower end of the specimen was held in a universal joint, and the stress applied to the upper end by a flexible wire, every care being taken to ensure conditions of axial loading. The apparatus was designed to enable the specimen to be surrounded by a copper can which in turn could be immersed in a constant temperature bath.

The light-weight extensometer (11 gm.) developed for this work was of the optical lever type, and measured the extension of a 5 cm. gauge length. Figure 2

* The small scale chemical analysis necessary to reveal this was kindly undertaken for us by Messrs. Thos. Bolton and Sons, to whom our thanks are due for this service.

shows a drawing of the extensometer mounted on a specimen, and with the two optical levers removed for clarity: one of them is sketched alongside. The gauge length was defined by the knife-edge V's (K_1 and K_2) each carried by a clip held firmly to the specimen by a light steel spring: one such is shown just above K_1 . The steadying V's (V_1 and V_2) rested only lightly on the specimen, being cut in thin sheet steel springs S_1 and S_2 which needed only a very small force to move them parallel to the specimen axis. The lower clip carried two vertical comparator rods, C, C', attached to two adjustment screws, F, F', by short lengths of steel wire, H, H'. The four needle points on the upper clip were in a line intersecting the specimen axis. The needle points, X, Y, Z, supported one of the optical levers; the remaining three supported the other. The optical levers were located uniquely on the needle points by means of two conical depressions and one V-groove punched in each lever base, firm contact being ensured by light springs T, T', which coupled on to hooks, R, R'. The flexible couplings H, H', provided the necessary freedom for the needle points, X, X' (which were located in conical depressions) thus avoiding over-constraint of the optical levers.

The two optical lever mirrors were inclined at about 45° to the horizontal, light entering and leaving the system vertically. The rest of the optical system consisted of an illuminated slit in the focal plane of a long focus lens ($f=120$ cm.); the image, formed in the same plane after reflection from the two mirrors in turn, was observed with a travelling microscope. The instrument could measure strains up to 10^{-2} without re-setting, and the accuracy of the values of strain was about 10^{-6} .

This arrangement had the following advantages: (i) errors due to end effects were eliminated by measuring over a central gauge length; (ii) thermal expansion of the specimen was approximately compensated by that of the comparator rods, made of similar material; (iii) use of the double mirror system not only doubled the sensitivity, but eliminated all errors due to changes in orientation of the specimen relative to the tensile machine; (iv) since the measurements were virtually made on the specimen axis, errors due to straightening of a specimen initially slightly bent were minimized.

In order to get a high rate of recording, an aircraft gun camera, adapted to take single exposures, was used to photograph simultaneously coarse and fine scales coupled to the travelling microscope, and a stop watch. The camera was electrically operated by push-button control each time the microscope cross-wire and the image were made to coincide. Between 20 and 30 accurate readings per minute could be obtained with this arrangement.

In some cases of transient creep, the initial creep rate was too great to enable the travelling microscope to follow the image and the first reading was only possible about a minute after commencement of creep. To obtain a record of this rapid creep, when the image was displaced several centimetres, a length of sensitive 35 mm. film was placed to receive the moving image. A rotating disc with cut-out sectors, driven by a synchronous motor, was placed in front of the illuminated slit. The interruptions served as time markers, so that 1, $1/4$, or $1/10$ of a revolution could be distinguished. The period of rotation was 1.1 seconds: thus intervals as short as 0.11 second were recorded. When the image displacement became slow enough for microscope observation, the film holder could be swung out of position and microscope settings made in the ordinary way.

§ 4. RESULTS

(a) Creep Tests on Virgin Specimens

Preliminary observations of strain made with a slowly increasing stress applied to a specimen showed that the material had a fairly well defined yield point. When the stress was very nearly equal to the yield stress, the creep behaviour was complex; the creep rate after falling from its initial value rose to a maximum before finally decaying to zero. This behaviour was only shown over a narrow range of applied stress. Further experimental work to elucidate the mechanism is in progress and it will not be considered further here.

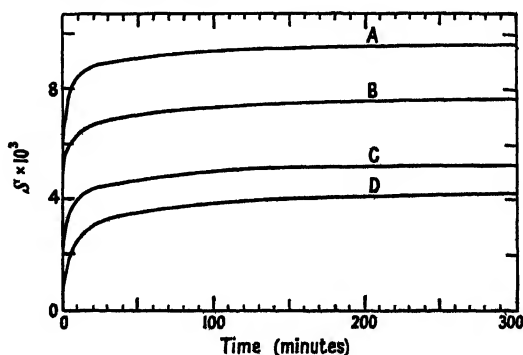


Figure 3. Creep of polycrystalline wires at room temperature. Stresses in kg/mm^2 : A, 6.55; B, 6.00; C, 5.77; D, 5.60.

At rather higher stresses the behaviour was normal. Figure 3 shows a number of graphs of strain plotted against time for polycrystalline specimens at room temperature. Each curve refers to a wire which had not previously been subject to any plastic deformation and which was loaded to the stress shown in the figure. Only the early parts of the curves are plotted, although the observations were usually

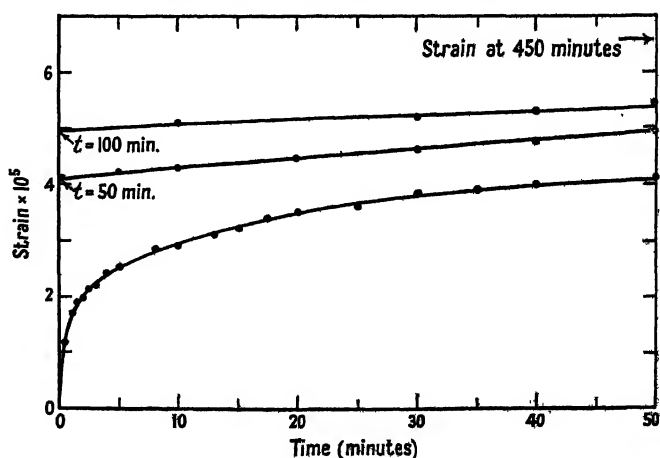


Figure 4. Creep at 90°A. ; stress = 9.50 kg/mm^2 .

continued for at least 24 hours. The creep rate decreased continuously during this period. Figure 4 shows the early portion of another similar creep curve on an

enlarged scale; this illustrates the accuracy with which the observed points lie on a smooth curve. Many more observations were in fact available than are plotted even on this graph.

The curves of Figure 3 are of the typical transient creep type.* Similarly shaped curves were obtained for wires extended at liquid oxygen temperature, although the stress level was about twice as great. In order to see if the observations could be fitted to an equation similar to (5), graphs were plotted of strain to the power $3/2$ against the logarithm of time; the result should be a straight line. Figure 5 shows a typical set of curves. These are straight for large times ($t > 10$ minutes) but show a curvature in the early stages, being concave towards the axis of time. This method of plotting shows up any disagreement between experiment and theory more clearly than a simple linear plot of (strain, time) in which experimental points are compared with a fitted theoretical curve. By extrapolating the straight portion one can obtain values of the constants ν and $K\{\nu N_0 \sigma_0 (kT/B)^{2/3}\}$ of equation (5).

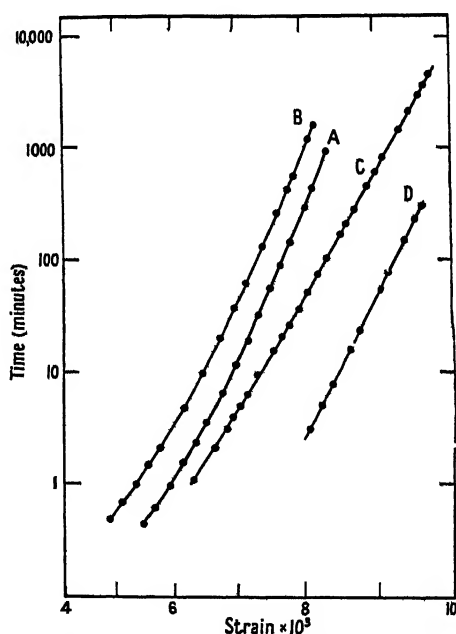


Figure 5. Creep of polycrystalline wires, A, B and D at room temperatures, C at 90° A. Stresses in kg/mm^2 : A, 5.70; B, 6.00; D, 6.55; C, 10.2.

The values of ν found were of the order of 10^2 to 10^4 per second, whereas the value expected from theoretical considerations is more nearly 10^9 per second. There is reason to believe, however, that the discrepancy is even greater than this would indicate. The theory predicts an 'instantaneous' strain occurring immediately the stress is applied, followed by the transient creep, the initial creep rate being very large but decreasing very rapidly. Equation (5) only attempts to account for the transient component of this creep. Now, in practice it is impossible to determine where instantaneous strain ends and transient creep begins, even with the photographic film method of recording mentioned in §3, because a finite time is required to apply the stress, and the extensometer has a finite response time.

* All strains, as plotted, are plastic components; elastic components have been subtracted.

Consequently, the strains as measured should be reduced by an unknown amount to allow for the instantaneous strain, before the experimental data are compared with theory. The effect of this is two-fold: firstly, the curvature of the graphs of the type shown in Figure 5 is made somewhat more pronounced*, and secondly, the values of ν obtained by extrapolating the linear portion are reduced even further. It is noteworthy in this connection that those experiments giving the highest values for ν were those where the stress was highest and therefore the error due to the inclusion of the instantaneous strain was also greatest.

This difficulty is sometimes avoided by treating the instantaneous strain as an extra disposable constant in fitting an equation to the experimental data (cf. Andrade 1910, 1914). In the present instance, in view of the good accuracy of the

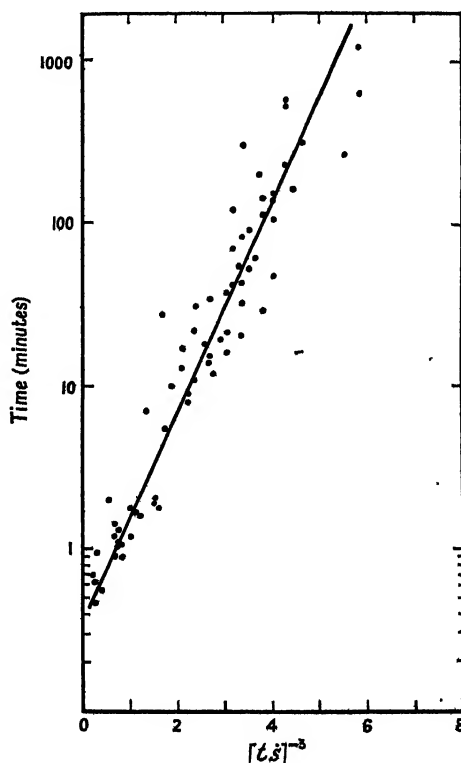


Figure 6. Creep at room temperature; stress = 6.00 kg/mm². Strain in arbitrary units. The data are the same as those represented by curve B of Figure 5.

observations of strain, it seemed preferable to analyse the results in terms of the strain-rate rather than strain, whereupon the difficulty disappears. By differentiating equation (5) one obtains immediately,

$$(\dot{\epsilon}t)^{-3} = \text{const.} (\log \nu + \log t),$$

which suggests plotting $(\dot{\epsilon}t)^{-3}$ against $\log t$. The errors involved in determining the creep rate are, of course, magnified when $(\dot{\epsilon}t)$ is raised to the third power, and thus the points show a considerable scatter. Figure 6 shows a typical curve.

* The linearity can be improved by plotting the strain to some higher power than 3/2, but this cannot reasonably be done within the framework of the Mott-Nabarro theory, since it would require $m < 0$ in equation (6) and hence $N = \infty$ when $\sigma = \sigma_0$.

Estimating a 'best' straight line through these points gives a value for ν of about 1 per second.

It is interesting to note in passing, that this method of analysing the results gives a sensitive method of distinguishing between the Mott-Nabarro equation and the Andrade equation (Andrade 1910, 1914). Under the conditions of these measurements, Andrade's ' k -flow' is likely to be negligibly small, and his equation thus reduces to :

$$L = L_0(1 + \beta t^{1/3}),$$

i.e.

$$S = \beta t^{1/3} \text{ for small strains.}$$

On differentiating this gives :

$$(t\dot{S})^{-3} = (3/\beta)^3(1/t).$$

Thus if $(t\dot{S})^{-3}$ is plotted against t , the Andrade equation would predict a curve decreasing hyperbolically, while the Mott-Nabarro equation predicts a curve increasing logarithmically. The results obtained in this investigation are unmistakably of this latter rather than the former type. The behaviour of this alloy is thus quite different from that of many pure metals which have given transient creep curves in agreement with the Andrade equation.

(b) Incremental Loading Creep Tests

A number of creep tests were carried out in which the loading programme was as follows. The specimen was first loaded to a stress at, or just above, the macroscopic yield point, and the creep allowed to proceed for about 24 hours. By this time the creep rate was very small. A stress increment was then made and creep readings taken over a period of about another day. Sometimes two successive increments could be made on the same specimen. The behaviour appeared to depend on the magnitude of the stress increment rather than on the absolute stress level. It was found that the behaviour could be reasonably accounted for in terms of the concepts of the Mott-Nabarro exhaustion theory of creep and the results obtained will be discussed in terms of that theory as outlined in § 2.

The application of a first stress σ_0 will use up all the most easily movable dislocations and, after a time t_1 the step discussed in § 2 will advance along the axis of σ to a point σ_1 , say (see Figure 1). Almost all the dislocations lying to the left of σ_1 will have moved under the applied stress, with the aid of temperature fluctuations, while those to the right of σ_1 will be only little affected. The actual advance $(\sigma_1 - \sigma_0)$ will depend on the nature of the material, the temperature, and the time. It has already been shown that the advance is approximately proportional to $(\log \nu t)^{2/3}$, so that after 24 hours (as in this experiment) it will be moving only slowly.

The material might now be expected to show fairly marked yielding for increments of stress greater than $(\sigma_1 - \sigma_0)$ with an 'instantaneous' plastic strain—as in the case of a virgin specimen stressed above its yield point—followed by a rapid, but rapidly decreasing creep. If the stress increment is less than $(\sigma_1 - \sigma_0)$, however, there should be no instantaneous plastic strain, and the initial creep rate should be much lower.

It was to obtain evidence upon this point that the method of photographic recording, described at the end of § 2, was devised. The stress increments were applied as quickly as possible without causing any jerks, but, even so, observations in the first half-second are probably not reliable. The results obtained for three

different stress increments are shown in Figure 7. It will be seen that two of them ($\Delta\sigma = 1.60$ and $\Delta\sigma = 1.86 \text{ kg/mm}^2$) give strain curves showing a very rapid initial strain followed by a typical transient creep curve, whereas the other

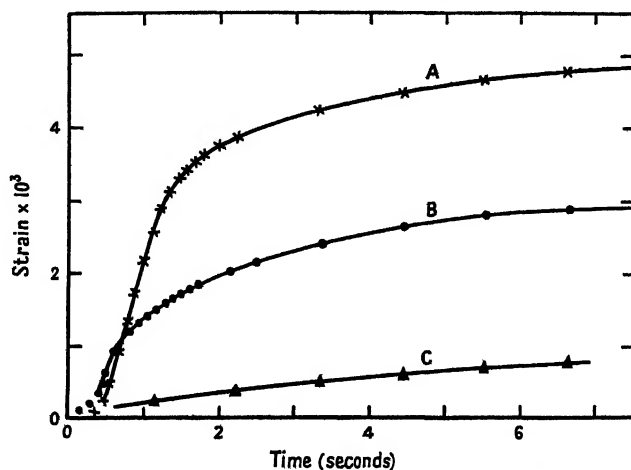


Figure 7. Initial parts of creep curves for incremental loading, room temperature. Stress increments in kg/mm^2 : A, 1.86; B, 1.60; C, 1.20.

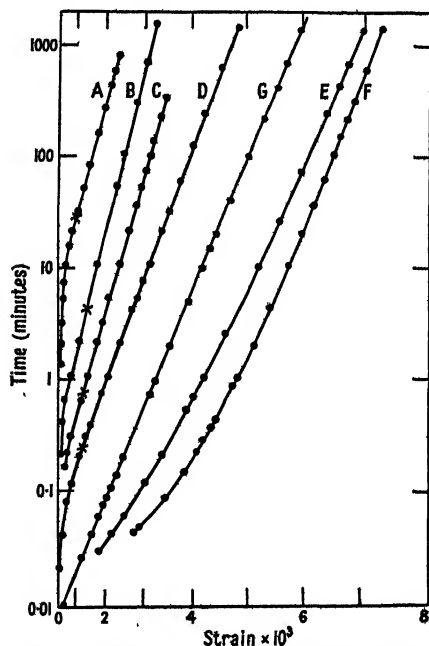


Figure 8. Creep curves for incremental loading, room temperature. Stress increments in kg/mm^2 : A, 0.10; B, 0.45; C, 1.01; D, 1.20; E, 1.60; F, 1.80; G, 1.40.

($\Delta\sigma = 1.20 \text{ kg/mm}^2$) has quite a small initial creep rate. From a series of such observations, it was concluded that the 'yield point' for such incremental loadings was at about 1.4 kg/mm^2 when the previous stress had been applied for about 24 hours.

The results of such a series of tests are shown in Figure 8, in which strain to the power $3/2$ is plotted against $\log(\text{time})$ as before. For stress increments greater than the critical value the shape of the curves is the same as that obtained with a

virgin specimen—concave towards the time axis initially and tending to become linear for large times. For increments less than the critical value the curvature has the opposite sign, while when $\Delta\sigma = 1.40 \text{ kg/mm}^2$ the points lie on a straight line.

When the increment is less than the critical value it is instructive to plot the experimental data in another way. We recall that Mott and Nabarro adopt a function, $N = N_0(1 - \sigma_0/\sigma)^m$ to describe the distribution of dislocations in terms of the stress σ . A large m corresponds to a curve rising gradually from zero at $\sigma = \sigma_0$, while $m = 0$ corresponds to $N = N_0$, a constant, for all σ . Making this assumption, equation (5) becomes

$$S = \text{const.} (\log vt)^{2(m+1)/3}. \quad \dots\dots (7)$$

Differentiating (7), and then eliminating v we obtain

$$\log (t\dot{S}) = \text{const.} + C \log S, \quad \dots\dots (8)$$

where

$$C = (2m - 1)/(2m + 2).$$

Thus a graph of $\log (t\dot{S})$ against $\log S$ should give a straight line of slope C . Note that in applying this analysis to the incremental creep curves under consideration, there is no ambiguity in the value of the strain S since there is no instantaneous strain. The limiting values of C are: $C = -\frac{1}{2}$ for $m = 0$ and $C = +1$ for $m \rightarrow \infty$.

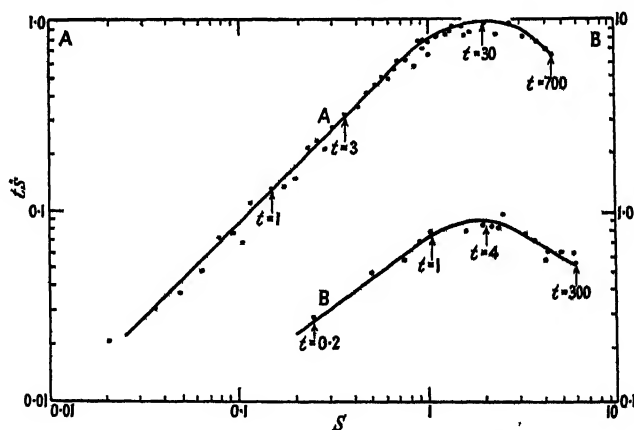


Figure 9. Incremental creep at room temperature. Stress increments in kg/mm^2 : A, 0.10; B, 0.64. Strain in units of 5.4×10^{-4} .

Figure 9 shows an example of the results obtained when $(t\dot{S})$ is plotted against S on a logarithmic scale for two incremental creep experiments. In the early stages the graphs are linear with a slope C approximately equal to 1; subsequently the curve passes through a maximum, and we have eventually a slope given by $-1 < C < 0$. In view of the scatter of the points available, it is not possible to determine the value of the negative slope at all accurately, but the data are not inconsistent with its being $-\frac{1}{2}$.

The immediate conclusion is that no simple law, such as equation (6), can describe the distribution function $N(\sigma)$ as it existed before the stress increment was applied. We would indeed expect it to be largely determined by the form of the step function $\exp(-\alpha t)$ corresponding to the initial stress. The problem of calculating the creep curve with a distribution function of this type appears to be quite intractable. One point may, however, be made. The treatment of exhaustion

creep theory given in §2 shows that the creep rate at any time is determined almost entirely by the shape of the distribution curve near the value of σ to which the step has advanced, provided that $N(\sigma)$ is not varying rapidly with σ at this point. In the incremental loading tests under discussion, the step will eventually advance beyond the escarpment in the distribution function into the region where $N(\sigma)$ will not vary rapidly. The time taken to reach this situation will be less when the increment is greater, and will be zero when the increment equals to $(\sigma_1 - \sigma_0)$. It is suggested that the maximum in the graph of $\log(t\dot{S})$ against $\log S$ corresponds to the transition from a regime in which $N(\sigma)$ is varying rapidly, to a regime in which $N(\sigma)$ is much more nearly constant. It is in fact observed that this transition takes place earlier when the stress increment is greater. This is shown by the data given in the Table.

Specimen	$\Delta\sigma$ (kg/mm ²)	t for $(t\dot{S})_{\max}$ (minutes)	S at $(t\dot{S})_{\max}$ ($\times 10^8$)
P 23 F	0.10	30	1.05
P 24 C	0.37	25	1.0
P 24 B	0.45	7	1.2
P 23 C	0.64	4	1.15
P 31 B	1.01	$\frac{1}{2}$ to 1	1.2

One would also expect that the total creep strain up to this point would be the same whatever the stress increment, since it corresponds to using up all the dislocations up to the escarpment on the distribution curve. This is fairly well borne out in practice, as the last column of the Table shows the strain at $(t\dot{S})_{\max}$ to be reasonably constant.

Only when this maximum has been passed would one expect the Mott-Nabarro creep equation to be valid with the index $2/3$; in Figure 8 we observe that the relevant curves (A to D) become straight lines only for strains greater than those corresponding to the maximum values of $(t\dot{S})$ —marked in Figure 8 by crosses.

When the stress increment is chosen to be just equal to the critical value of 1.4 kg/mm. one might expect the conditions to be most favourable for obtaining agreement with equation (5). It will in fact be seen (Figure 8, curve G) that the linear relation between $\log t$ and $S^{3/2}$ holds over the entire range from 10^{-2} to 10^8 minutes. By extrapolation of this line, one can obtain the value of ν more reliably than heretofore; the result is 10^2 /minute.

§ 5. DISCUSSION

It will be clear from the previous section that the effects observed in the experiments on incremental loading are all readily accounted for in a qualitative manner in terms of exhaustion creep theory. A more quantitative treatment is difficult and it is doubtful if the effort would be justified at this juncture. The general form of the strain-time relation is also consistent with the theory.

When we consider the quantitative results however, the position is less satisfactory. The surprisingly low value of ν —the frequency of vibration of a dislocation loop—has already been remarked. If however, we accept this result, then a knowledge of the distance that the 'step' advances along the axis of stress in a given time enables us, with the aid of equation (4), to evaluate the quantity B of equation (3). Taking the advance $(\sigma_1 - \sigma_0)$ to be 1.4 kg/mm² in 24 hours, one finds that B is of the order of $40kT$ —a not unreasonable figure. Further, from equation (1) setting $\sigma = 8$ km/gm²—the order of magnitude of the stresses used—we

find that Λ , the length of a dislocation loop, is about 60 Å: this too is not impossible, although smaller than might have been expected. Finally, the slope of a graph of $S^{3/2}$ against $\ln(vt)$ then enables us (equation (5)) to obtain a numerical value for vN_0 . The incremental loading curve for $\Delta\sigma = 1.4 \text{ kg/mm}^2$ is again the most suitable one to choose, and gives $vN_0 = 1.7 \times 10^{-11}$.

We cannot evaluate either v or N_0 separately without some further assumption and any further discussion must be somewhat speculative. It will be recalled that $Nd\sigma$ was defined as the number of dislocation loops per unit volume with activation stresses in the range $d\sigma$. Now the total number of dislocation lines per unit area in a well annealed crystal has been estimated as being of the order of 10^8 . Accepting this value for the moment, we obtain

$$\int_0^\infty Nd\sigma = 10^8/\Lambda;$$

using the value of Λ already found, this is of the order $10^{14}/\text{cm}^3$. On our assumption that N is independent of σ , this must be set equal to $N_0(\sigma_{\max} - \sigma_{\min})$ where σ_{\min} is the least applied stress that will move a dislocation loop (of the order of the yield stress), and σ_{\max} is the stress required to move the most tightly bound dislocation loop. σ_{\max} must exist, if only for the reason that the material will not support infinite internal stresses; let us suppose that it lies between 2 and 11 times the yield stress. Then $(\sigma_{\max} - \sigma_{\min}) = \sigma_{\min}$ (or $10\sigma_{\min}$ on the second assumption), i.e. is of the order 10^9 (10^{10}) dyne/cm². Thus $N_0 = 10^5$ (10^4) and this would suggest $v = 10^{-16}$ (10^{-15}) cm³. Now v/V was defined as the strain resulting from the activation of a single dislocation loop, where V is the volume of the crystal. The movement of the activated loop itself is equivalent to slip of one atomic spacing over an area of slip-plane of order of magnitude Λ^2 , and thus corresponds to a strain of $a\Lambda^2/V$. Thus if the activated loop is the only one that moves, we have $v = a\Lambda^3$, which is of order 10^{-20} cm³. Comparing the two figures we may deduce that the activation of a single dislocation loop starts off an avalanche which, on the average, involves about 10^4 (10^5) other similar loops.

Although the preceding calculation is very rough it does suggest that the elementary slip process with which the theory ought to concern itself is one involving a considerable number of dislocation loops. It may be that this will provide a clue to the major problem, that of reconciling the theoretical and experimental estimates of the frequency of vibration of a dislocation loop. Until this is done the theory cannot be considered to be entirely satisfactory. It is also clearly desirable to extend the measurements to other 'hard' materials, to which the theory should apply equally well; it is hoped to do this in due course.

ACKNOWLEDGMENTS

Our best thanks are due to Professor Mott for suggesting this problem and to many colleagues in the H. H. Wills Physical Laboratory for assistance with the work and discussion of the results. One of us (M. D.) also wishes to acknowledge financial assistance from the Ministry of Education under the Armed Forces Further Education Scheme.

REFERENCES

- ANDRADE, E. N. DA C., 1910, *Proc. Roy. Soc. A*, **84**, 1; 1914, *Ibid.*, **90**, 329.
 MOTT, N. F., and NABARRO, F. R. N., 1948, *Report on Bristol Conference on Strength of Solids* (London: Physical Society), p. 1.
 TAYLOR, G. I., 1934, *Proc. Roy. Soc. A*, **145**, 362.

A Study of Electrical Forming Phenomena at Selenium Contacts

By H. K. HENISCH AND J. EWELS

Physics Department, University of Reading

Communicated by R. W. Ditchburn; MS. received 13th January 1950

ABSTRACT. An account is given of current-creep experiments at various temperatures on Se specimens of different impurity content. The dependence of these phenomena on the electrical and thermal history of the specimens and on the nature of the counter-electrode is investigated. It is shown that two opposing creep mechanisms are, in general, active simultaneously. One is due to power dissipation within the barrier layer, the other due to structural changes which take place under the influence of the applied field. It was found, contrary to theoretical expectations, that creep processes cannot be 'frozen out' at low temperatures (e.g. -183°C.). Measurements of the self-capacitance of the barrier were carried out at suitable stages of the experiment. These show that the thickness of the barrier increases slightly during forming. A theory of current creep is proposed on the basis of the present observations.

§1. INTRODUCTION

WHEN a reverse voltage is applied to a rectifying contact between a selenium surface and a metal electrode, the resulting current is not constant but is generally a function of time. An increase of current with time is called positive creep and a decrease at constant voltage is referred to as negative creep. These phenomena, which can also be observed on other semiconducting materials, are the basis of the forming process which is of great practical importance in the manufacture of disc rectifiers, crystal diodes and crystal triodes (transistors). The forming process as applied to selenium rectifiers results in a very considerable improvement of the rectification properties. Junctions which have been 'electrically formed' have a higher D.C. resistance in the blocking direction and are therefore able to withstand a much higher reverse voltage without breakdown than unformed discs. It must be inferred that some structural modification of the barrier layer takes place under the influence of the applied field which may be as high as 10^6 volts/cm. At normal temperatures a formed rectifier retains its improved characteristics, except for the slow deterioration known as ageing. The structural modification of the barrier must therefore be ascribed to a process which is only slightly reversible under these conditions. Its exact mechanism is evidently of great interest, not only because of its technical importance, but because no theory of barrier layer rectification can be complete unless it can account for creep phenomena.

A good deal of information on these matters is probably in the hands of rectifier manufacturers, but very little has appeared in the technical press. Williams and Thompson (1941) give a brief description of forming phenomena on commercial rectifiers. Their results show the complicated nature of the processes involved and their dependence on the thermal and electrical history of the specimens. Rose and Schmidt (1947) have studied similar phenomena on selenium contacts by the application of voltage pulses lasting about one-fifth of a second, and claim to have shown that positive creep is due to an increasing concentration of impurity centres in the immediate neighbourhood of the counter-electrode, and negative creep due

to the subsequent migration of centres from the semiconductor into the metal under the influence of the applied field. More recently, Kobayashi (1949) has published a short account of creep phenomena in cuprous oxide rectifiers. In the course of his investigations, the time constant of negative creep was found to be about 2 minutes.

The purpose of the experiments described in the present paper was to obtain a more comprehensive series of results, not on commercial rectifiers whose structure is rather complicated, but on selenium contacts prepared under controlled laboratory conditions*. Amorphous selenium, either pure or with deliberate additives, was deposited on nickel-plated steel discs, as during the manufacture of normal rectifiers, and subsequently heat-treated to convert it to the crystalline form. The counter-electrodes used were either gold (applied by vacuum evaporation) or a low melting-point alloy of 72 % Sn and 28 % Cd (applied by spraying). The experiments consisted mainly in the continuous observation of the currents resulting from the application of various constant reverse voltages to selenium contacts. At convenient intervals the self-capacitance of the contacts at zero bias voltage was determined by measuring the resistive and reactive components of the contact impedance at different frequencies, as described by Henisch (1949). The influence of temperature, electrode material, interruptions in the forming process and impurity content of the selenium on the capacitance and the current-time relation was also studied. The systematic variation of these parameters has led to results by which existing theories of current creep can be tested. It is also possible to obtain a clearer picture of the conditions within the barrier, particularly as regards the relations between thermal effects and ionic diffusion.

§ 2. EXPERIMENTAL TECHNIQUE

The simple circuit used for the present investigation is shown in Figure 1. The reverse voltages applied to the contacts under test were taken from a series of 2 v. accumulators. A potential difference developed across the input coil of the magnetic amplifier and the resulting output current was passed to a recording

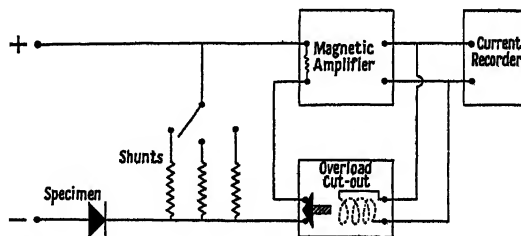


Figure 1. Circuit for current-creep measurements.

milliammeter (5 ma. full scale deflection, chart speed 1 in. per min.). The input resistance of the amplifier (6.61 ohms) could be shunted so as to provide various current ranges (e.g. 0–6 ma., 0–30 ma., 0–120 ma., 0–1 amp.). A valve-controlled overload cut-out was installed to protect the current recorder in case of electrical breakdown of the specimen under test. The accumulators were periodically charged *in situ* and before use partly discharged through a dummy load in order to ensure constancy of the applied voltage. Owing to the inertia of the current

* Supplied by Messrs. Standard Telecommunication Laboratories Ltd. to the authors' specification.

recorder no observations were possible during the first second or so after application of the voltage.

The bridge circuit used for capacitance measurements is shown in Figure 2. The decades in the balancing arm could be arranged either as a series or a parallel combination, whichever proved more convenient for a particular specimen. The applied alternating voltages never exceeded 20 mv.

The specimens consisted of discs of 3.5 cm. diameter which carried six independent electrodes as shown in Figure 3. One or other of these contacts could be brought into the circuit of Figure 1 by means of a selector switch. During the experiments the discs were mounted in a tubular brass container which could be evacuated and heated (electrically) or cooled (for instance, by means of solid CO_2) from the outside. The ambient temperature was measured by means of a chromel-alumel thermocouple at the centre of the disc. The temperature of the counter-electrode may be somewhat higher, depending on the power dissipated within the

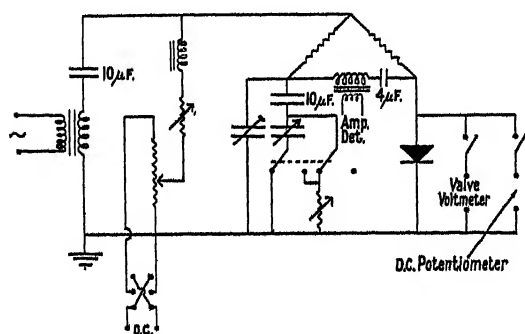


Figure 2. Bridge circuit for the measurement of barrier self-capacitance.

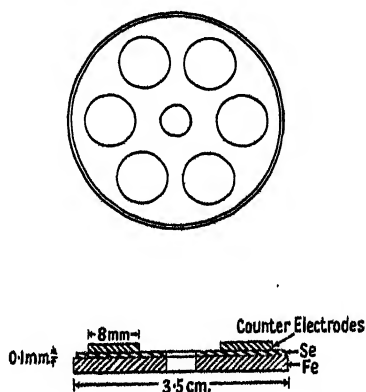


Figure 3. Experimental specimens.

barrier layer. During one series of experiments the resulting temperature rise was approximately assessed by placing a thermojunction in direct contact with the counter-electrode. All current leads were taken out of the container through hermetic seals which were electrically heated to prevent frosting over during the low temperature experiments.

§ 3. GENERAL CHARACTER OF CURRENT-CREEP PHENOMENA

The general creep character of the observed current creep is shown in Figure 4 (curve *a*). The application of a constant reverse voltage to the contact is followed by a short period of positive creep and, in due course, by a much longer period of negative creep. It is reasonable to regard a curve of this kind as the composition of two opposing tendencies, representing pure positive creep (curve *b*), and pure negative creep (curve *c*) respectively. There is no physical reason for expecting these components to be simply exponential. The usefulness of this analysis depends ultimately on the possibility of showing that two distinct—though not necessarily wholly independent—physical processes are at work. Evidence for this conclusion will be given and discussed in the following sections.

In Figure 4, the speeds of the two processes and their amplitudes are assumed to be comparable. In general this need not be so. Hence, while opposing tendencies may both be operative, the current maximum may or may not be

observed, depending on the relative magnitude of j_H and j_F as well as the speeds of the corresponding processes. The current-time relation may thus appear in a number of forms, examples of which are shown in Figure 5. Positive or negative creep may be suppressed or, in a special case, the two processes may balance initially (e.g. as in curve *c*) so as to keep the current temporarily constant. Simple tests show that curves of type *a* and *b* are indeed not simply exponential and it is

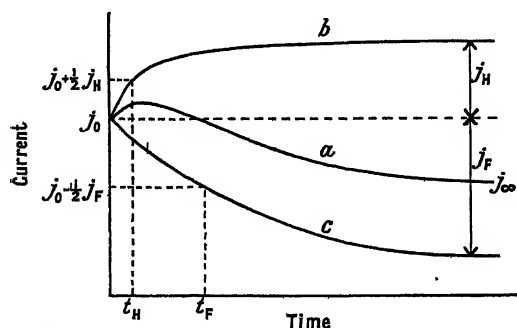


Figure 4. General character of current creep phenomena (schematic).

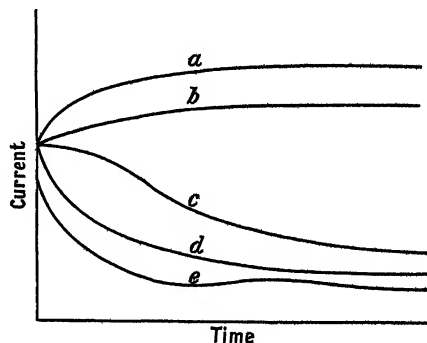


Figure 5. Some special cases of current creep (schematic).

thus impossible to define time constants with precise theoretical meaning. However, the 'periods of semi-completion' (t_H and t_F) can be taken as rough measures of the speeds. A review of the results shows that t_H is usually smaller than t_F .

The above comments relate to changes in the reverse currents with time. No current creep of any kind was observed for constant voltages applied in the forward direction.

§ 4. INFLUENCE OF ELECTRICAL HISTORY

For the sake of simplicity and in order to establish a basis for comparison, the discussion in this and the two following sections will be restricted to experimental results obtained on discs with alloy counter-electrodes on selenium containing 0.02 % of iodine. In order to assess the influence of electrical history, five types of current-creep experiments were carried out:

- (a) formation of a rectifying junction at successively increasing reverse voltages up to a safe maximum, insufficient to cause breakdown;
- (b) formation of a rectifying junction at a single high reverse voltage;
- (c) application of a low reverse voltage shortly after formation of the barrier at a higher voltage;
- (d) repeated application of the highest previous forming voltage at short intervals;
- (e) repeated interruption of the forming process before the establishment of equilibrium, i.e. before a constant current has been reached.

Figure 6 shows the progress of current creep on two new junctions at reverse voltages increasing in steps of about 4 volts, with 10-minute intervals between successive stages. Capacitance measurements were carried out before this sequence and during the intervals. In order to save space, only the first six minutes of the creep record are reproduced. Where applicable, the current at the conclusion of the run and the corresponding time are indicated on the right-hand

side of the graphs. It should be noted that the scale and origin on the current axis is not the same for all curves. Negative creep is observed in every case, but the initial positive creep (and hence the current maximum) only for higher voltages. Different electrodes on the same disc behave in an essentially similar way, though the values of j_0 and j_∞ may vary somewhat from one junction to another.

These results are typical for current creep at room temperature. Experiments carried out at different temperatures—and, of course, on different junctions—are not strictly comparable in view of the normal variations between different electrodes, and because the effect of temperature changes is not perfectly reversible. Even in absence of an externally applied voltage, the barrier carries the normal diffusion field

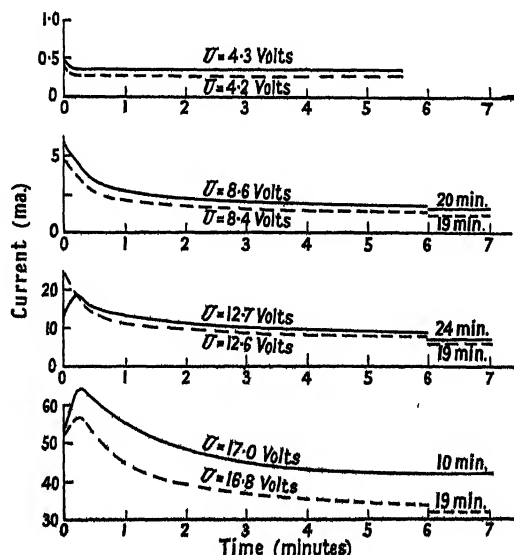


Figure 6. Record of creep phenomena during forming at successively increasing reverse voltages (U).

Se+0.02% I_2 alloy counter-electrodes.

10 minute intervals between runs.

— Disc b, electrode 1, ambient temperature 22° c.

--- Disc b, electrode 2, ambient temperature 19° c.

under the influence of which almost irreversible modifications in the structure of the barrier may take place. Experiments carried out at low temperatures (e.g. -59°C .) and high temperatures (e.g. $+60^\circ\text{C}$.) do not reveal any characteristic deviations from the relations shown in Figure 6. Creep phenomena are still very pronounced at the lower temperatures and are not markedly enhanced at $+60^\circ\text{C}$., as compared with room temperature. Even at liquid air temperature negative creep is still clearly observable though no positive creep was encountered during the short experimental series in this range.

If, as an alternative to forming in stages, a high reverse voltage is directly applied to a new junction, the final resistances are not very different from those observed after forming in stages. They are reached after pronounced negative creep. The application of a comparatively low reverse voltage shortly after forming at a higher voltage results in positive creep only. An example of this is shown in Figure 7. The same is true if the highest previous forming voltage is repeatedly re-applied after short intervals, i.e. intervals not exceeding 1 hour or so. (Phenomena

observed after longer intervals will be dealt with in the next section.) Positive creep occurs in every case, to an extent which depends on the length of the preceding interval. Figure 8 shows how $j_{\infty} - j_0 (=j_H)$ varies with the length of this interval, for three different junctions. Such regular characteristics of successive creep curves on a given junction have never been observed in connection with negative creep. They are associated with positive creep only, a fact which supports

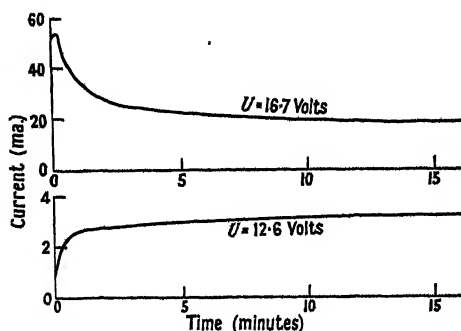


Figure 7. Current creep at a relatively low voltage after forming at a higher voltage. Se+0.02% I₂ alloy counter-electrode. 5 minute interval between runs. Disc e, electrode 4, ambient temperature 19.5° c.

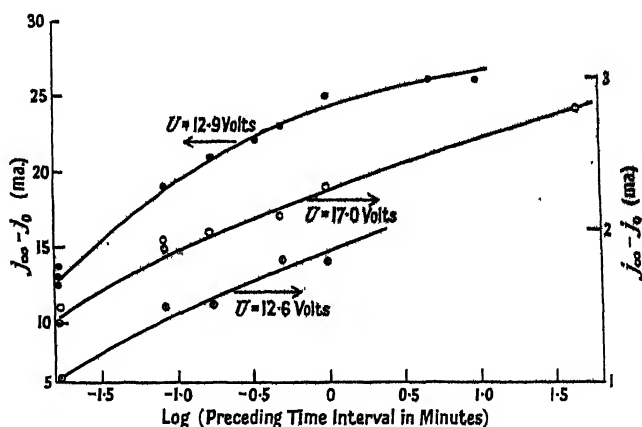


Figure 8. Current creep after re-application of the highest previous forming voltage. Relation between $j_{\infty} - j_0 (=j_H)$ and preceding time interval. Se+0.02% I₂ alloy counter-electrode.
 ○ Disc b, electrode 1, ambient temperature 22° c.
 ● Disc b, electrode 3, ambient temperature 17° c.
 ○ Disc c, electrode 2, ambient temperature -59° c.

the hypothesis that positive and negative creep are due to different mechanisms. This is confirmed by the form of the creep curve obtained when the forming process is repeatedly interrupted before equilibrium is reached, as shown in Figure 9. It will be noted that every short interruption (1 second or so) is followed by some positive creep, after which the original negative creep is resumed in much the same form as that in which it is known normally to proceed when the applied voltage is constant (broken line). If, on the other hand, the duration of the interruption becomes comparable with the total creep period, the current-time relation is as shown in Figure 10. The shape of the curve is the same as before, but a

displacement of the second part towards the right by an amount which corresponds to the period of interruption can now be observed. The broken line shows approximately the manner in which creep was expected to proceed if the applied voltage had remained constant. It appears that after the re-application of the external voltage, negative creep is resumed, but that its initial phases are more than compensated by positive creep.

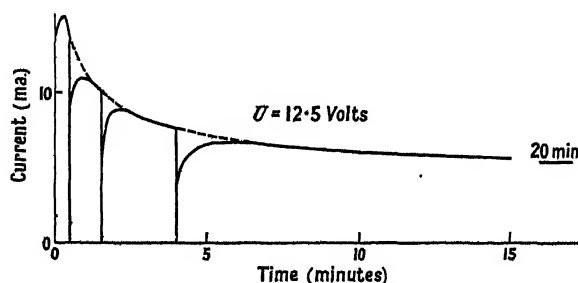


Figure 9. Current creep following short interruptions (1 sec.) of the forming process before the establishment of equilibrium.

Se+0.02% I_2 alloy counter-electrode.
Disc e, electrode 6, ambient temperature 21° c.

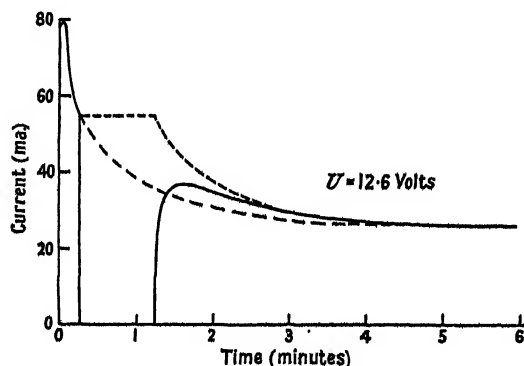


Figure 10. Current creep following a long interruption (approximately 1 minute) of the forming process before the establishment of equilibrium.

Se+0.02% I_2 alloy counter-electrode.
Disc g, electrode 2, ambient temperature 20° c.

§ 5. STABILITY

When a junction has undergone a series of electrical operations of the type shown in Figure 6, the barrier has not reached a perfectly stable condition. It is subject to further changes during storage, even in absence of an applied forming voltage. These changes are dependent on the storage temperature and can also be influenced by the passage of large direct currents in the forward direction. Both factors were investigated by observing original and final current-creep curves for a given applied voltage.

If a junction which has previously been formed at a high reverse voltage is stored for two days at room temperature and subsequently re-formed at the same voltage, the resulting creep curve is very similar to that observed before the storage period, but negative creep is less prominent. The final currents differ only slightly from the values of j_{∞} before storage. On the other hand, if similar measurements are carried out at a high temperature (e.g. 75 to 80° c.) the final currents are

considerably lower and are reached after an additional negative creep process. Figure 11 illustrates this point. Storage at high temperatures thus displaces the equilibrium previously reached at room temperature and allows a further modification of the barrier in the same direction.

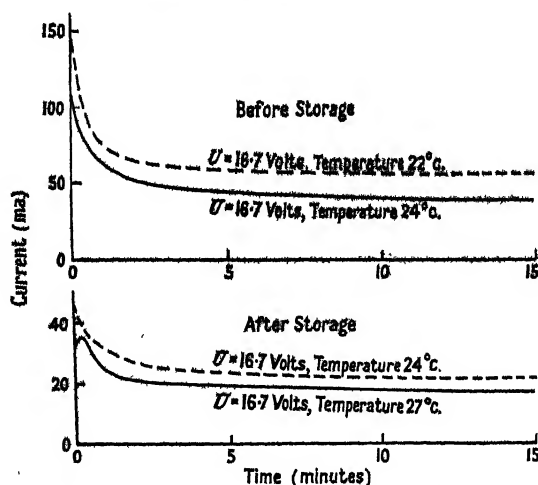
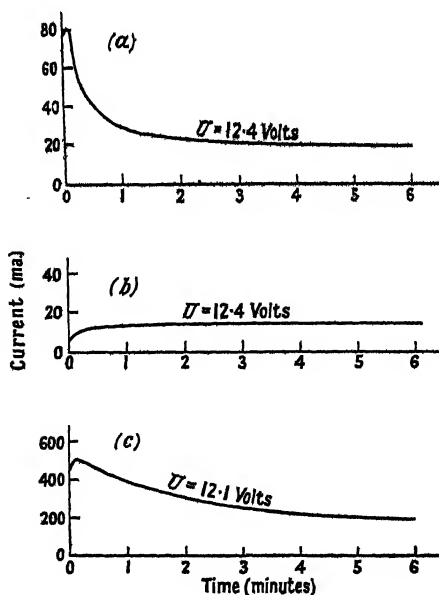


Figure 11. Current creep before and after storage for two days at a high temperature ($75-80^{\circ}\text{C}.$).

Se+0.02% I_2 alloy counter-electrode.

— Disc e, electrode 1.

--- Disc e, electrode 2.



(a) First application of forming voltage before passage of forward current.

(b) Re-application of forming voltage before passage of forward current.

(c) Re-application of forming voltage after passage of forward current.

Figure 12. Current creep before and after the passage of forward current (200 ma. for 45 hr.).

Se+0.02% I_2 alloy counter-electrode.

Disc g, electrode 6, ambient temperature $18^{\circ}\text{C}.$

The influence of forward current on the properties of the rectifying junction is shown in Figure 12. The first curve represents the original forming process at 12.4 volts. After an interval of 1 hour the same voltage was re-applied and the second curve shows the corresponding creep relation which is of the type already discussed in the previous chapter. A forward current of 200 ma. was then passed for 45 hours, and re-application of the original reverse voltage then resulted in current creep as shown by the third curve. It should be noted that j_{∞} has increased by a factor of about 10. It is clear that the prolonged passage of forward current causes a substantial modification of the barrier layer in the opposite direction to the change due to the forming process.

§ 6. VARIATION OF SELF-CAPACITANCE DURING FORMING

Schottky's well-known theoretical treatment of self-capacitance leads to the expression

$$\frac{d(1/c^2)}{dU} = \frac{8\pi}{KeN_{\lambda}},$$

where U is the applied voltage (direct), K the dielectric constant, e the electronic charge, N_{λ} the concentration of impurity centres at a distance λ from the metal boundary, and c the dynamic capacitance (dQ/dU) per unit area of the barrier layer. The above equation in conjunction with $c = K/4\pi\lambda$, can be used for a graphical evaluation of N_{λ} for any particular value of λ (Henisch 1949). This was thought to offer an attractive method of investigating the changes (if any) in the local concentration of impurity centres due to the various stages of forming. Such measurements would entail the determination of c for various applied voltages U , before and after forming. Since the application of these voltages to new junctions is itself responsible for the forming process, it was realized that Schottky's method could only be used if it were possible to 'freeze out' all creep phenomena by performing the capacitance measurements at a sufficiently low temperature. As already mentioned, forming phenomena are still observed at liquid air temperatures and the technique based on Schottky's capacitance relation could not, therefore, be applied as originally intended. In the cases here described, its usefulness was limited to the small range of reverse voltages (e.g. $U < 3$ volts) for which current creep is negligible. (In the forward direction the measurement of self-capacitance is more difficult owing to the low shunt resistance.) Figure 13 shows $1/c^2$ plotted against U . The resulting relation is a straight line which on the basis of Schottky's treatment implies a constant concentration of impurity centres. The value of N deduced from the slope is approximately 7×10^{16} . Some variations from one disc to another were observed, but this concentration can be taken as typical. The intercept on the abscissa should be numerically equal to the diffusion potential and is found to be about 0.6 volt. Unfortunately, the slope of such lines is not a sensitive measure of N , but the results indicate that forming produces a small increase in the concentration of impurity centres at distances between 2×10^{-5} cm. and 4×10^{-5} cm. from the counter-electrode.

Capacitance measurements in absence of a direct bias voltage are not, of course, subject to the above disturbances and represent a useful, though less informative, method of detecting changes in the structure of barrier layers. Such measurements were carried out after each stage of forming, and typical results are shown in Figure 14. It will be noted that the forming process is accompanied by a small decrease of self-capacitance, corresponding (as expected) to a thickening of the

barrier layer. The capacitances decrease with decreasing temperature. The series of measurements shown in Figure 14 were, of course, carried out on different junctions, but the self-capacitances determined at low temperatures were quite generally found to be smaller than those measured at high temperatures. The width of the barrier at normal temperatures is about 2×10^{-5} cm.

§ 7. EFFECT OF COUNTER-ELECTRODE MATERIAL AND OF ADDITIVES TO THE SELENIUM

Experiments of the kind described in the previous sections were also carried out on specimens with counter-electrodes of (vacuum-evaporated) gold, on selenium with and without deliberate additives. The results are summarized by Table 1 which gives the initial and final currents for two forming voltages.

Comparing values for j_0 , it will be seen that counter-electrodes of gold are always associated with lower barrier resistances than the corresponding alloy

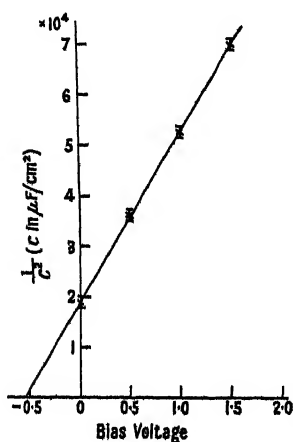


Figure 13. Schottky's capacitance relation plotted for disc e, electrode 5 after forming at 4 volts.

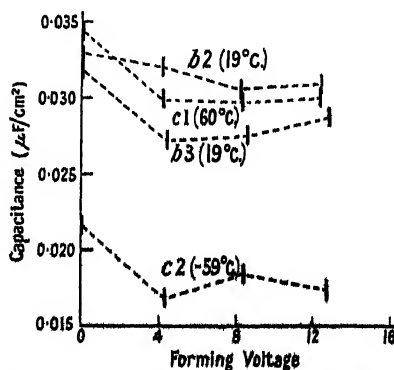


Figure 14. Changes of barrier capacitance due to various stages of forming. Se+0.02% I₂ alloy counter-electrode.

electrodes. This applies at both forming voltages. Table 2 gives the mean values of $(j_0 - j_\infty)/j_0$ (averaged over the whole temperature range) and the corresponding mean deviations for the various groups. These values show that the forming process is always less effective for gold than for the Sn-Cd alloy. The results also indicate that the effect of the additives to the selenium is remarkably small. This applies also to the current-time relation during forming, to the dependence of the forming process on the electrical history of the junction and to the self-capacitance. Iodine causes a small decrease of barrier resistance, and mercury (added before melting) a slight increase. Contrary to expectation, the initial and final currents of low-temperature runs are nearly always higher than those for room temperature and above.

The effect of mercury is quite different when specimens are exposed to mercury vapour. After short periods (e.g. several days in the presence of air at atmospheric pressure) the barrier resistance falls to a small fraction of its original value and the junction becomes useless for rectification purposes. This phenomenon represents a well-known danger to industrial rectification plant. The results

Table 1

Summary of Current Creep Experiments on Different Types of Junction

Counter-electrode	Additive	Junction No.*	Temp. (°C.)	Forming at 8.2 volts		Forming at 12.5 volts	
				j_0 (ma.)	j_∞ (ma.)	j_0 (ma.)	j_∞ (ma.)
Alloy	Nil	b1	+20	1.20	0.35	4.9	1.05
		b2	+22	0.65	0.30	4.0	1.20
		c2	+68	0.87	0.25	3.1	0.56
		c4	-54	6.2	1.2	27.0	4.2
		c5	-57	7.5	1.2	29.7	2.7
Gold	Nil	a2	+20	190	128	193	182
		a1	+23	109	99	—	—
		b2	+70	59	32	97	43
		b1	+70	80	42	—	—
		b5	-48	196	175	—	—
		b4	-50	262	210	—	—
Alloy	0.02% I ₂	b1	+22	4.9	1.6	16.0	7.0
		b2	+19	4.0	1.2	24.0	6.2
		b3	+17	4.2	2.6	22.0	7.5
		c1	+60	4.4	1.2	14.4	4.2
		c6	+60	1.6	0.8	5.0	3.1
		c3	-58	9.5	1.7	31.0	6.5
		c2	-59	9.2	1.7	42.0	6.6
Gold	0.02% I ₂	a3	+25	125	102	160	135
		a2	+30	140	109	—	—
		b2	+61	141	89	82	83
		b1	+61	170	95	—	—
		c2	-56	245	180	275	185
		c1	-57	38	(42†)	71	62
Alloy	0.05 %Hg‡	a3	+19	1.47	0.72	5.5	2.85
		a6	+65	0.78	0.39	3.3	0.97
		a5	+65	0.95	0.45	2.7	1.00
		a1	-49	9.6	2.30	35	5.5
		a2	-50	7.0	2.0	37	4.2
Gold	0.05% Hg‡	a2	+27	77	48	118	74
		a1	+22	83	53	129	71
		a6	+65	35	12.5	41	23.5
		a3	+60	55	30	88	44
		a4	-56	98	65	—	—
Alloy	0.02% I ₂	a2	+21	4.3	0.82	16.0	4.2
	0.05% Hg‡						

* The letters refer to different discs, the numbers to individual counter-electrodes (see Figure 3).

† Positive creep only.

‡ Added before melting.

previously quoted show that mercury is not harmful when added to the selenium before melting. After melting and crystallization it is then, presumably, present in interstitial positions or in some chemical combination, whereas exposure of the crystalline material to mercury vapour is more likely to result in adsorption at grain boundaries.

Table 2. Effectiveness of the Forming Process (at 8.2 volts) for Different Types of Junction

Counter-electrode	Additive	Average value of $(j_0 - j_\infty)/j_0$	Mean deviation
Alloy	Nil	0.72	0.08
Gold	Nil	0.31	0.14
Alloy	0.02% I ₂	0.66	0.13
Gold	0.02% I ₂	0.30	0.09
Alloy	0.05% Hg	0.60	0.11
Gold	0.05% Hg	0.43	0.09

§ 8. THE MECHANISM OF POSITIVE CREEP

The results described in the previous sections show that there are characteristic differences between the conditions for the occurrence of positive and negative creep. Thus negative creep is a process which depends critically on the previous history of the junctions and is generally not repeatable. Positive creep, on the other hand, is well reproducible when sufficiently high voltages are applied to junctions which are fully formed. It was therefore concluded that positive and negative creep are due to entirely different mechanisms, and that the observations of current creep under various conditions cannot be simply accounted for on the basis of existing theories.

The manner in which these differences manifest themselves indicates that positive creep is not caused by structural changes, but is due to the power dissipation within the barrier layer. This conclusion is supported by the fact that positive creep is always prominent when the power dissipation is high, and by the repeatability of the phenomenon. In order to examine this interpretation in more detail, a special series of measurements was carried out during which the temperature at the top of the counter-electrode (Figure 3) was repeatedly measured during negative and positive creep. Typical results are shown in Figure 15.

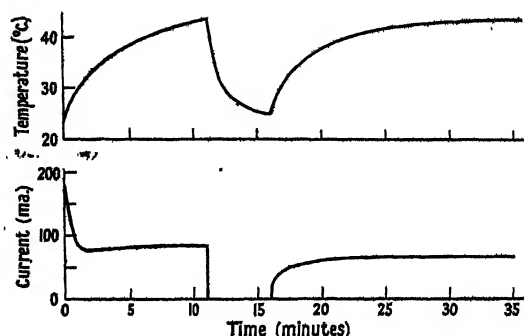


Figure 15. Record of counter-electrode temperature during an interrupted creep process.

Se+0.02% I₂ alloy counter-electrode.
Disc a, electrode 5.

The first application of a high reverse voltage results, as usual, in pronounced negative creep and also in some heating of the barrier layer. The actual temperature of the barrier layer itself is probably considerably higher than that given on the graph. This temperature rise counteracts the forming process and after

about two minutes (in this case) the two opposing tendencies balance. However, since temperature stability is not yet reached at this stage, the current thereafter increases, though more slowly. When the temperature of the barrier has become constant, slight negative creep may take over again until the forming process is completed. If the reverse voltage is then switched off, the counter-electrode cools down, as expected. Subsequent re-application of the same voltage results in prominent positive creep, followed by only slight negative creep, since the forming process was almost completed during the previous application. The corresponding temperature rise is again shown in Figure 15, and can always be observed whenever positive creep occurs. The second part of the graph in Figure 15 can be repeated many times without any appreciable change of shape.

On the above scheme one would expect that the re-application of a forming voltage after a long interruption should give rise to pronounced positive creep, whereas a short interruption during which there is little opportunity for cooling should be followed by only slight positive creep. This is confirmed by the results plotted in Figure 8.

The above interpretation implies that a rise in the temperature of the barrier necessarily results in a lowering of the barrier resistance. This has been confirmed by potential probe measurements on single barriers. The results in Table 1 show, however, that for the rectifier assembly as a whole, the resistance per unit area decreases as the temperature decreases. This behaviour has also been observed on previous occasions (U.S. Signal Corps Engineering Laboratories 1949) and its interpretation represents a major problem, independent of any considerations of rectifier forming. In some way the effect of heating the assembly as a whole is not the same as that of heating only the barrier layer and its immediate surroundings. Capacitance measurements have not revealed the existence of any resistance in series with the barrier which could be made responsible for the unexpected temperature dependence, even if a suitable mechanism could be postulated. It seems possible, therefore, that this behaviour is due to a change in the effective contact area with temperature. The manner in which such a change could be brought about is still uncertain. It may, conceivably, be due to differential expansion at the base electrode or in some other way due to lateral variations in the structure of the system. This matter will be referred to again in a future communication.

§ 9. THE MECHANISM OF NEGATIVE CREEP

An interpretation of negative creep must take into account the following observations:

- (a) Negative creep depends critically on the previous electrical and thermal history of the rectifying junction.
- (b) The effect of forming is only slightly reversible under normal storage conditions, but the prolonged passage of forward current results in a considerable amount of de-forming.
- (c) Storage of a fully formed junction at a high temperature (e.g. 75–80° c.) establishes conditions under which a further forming process is possible.
- (d) When allowance is made for an initial positive creep due to power dissipation, it is found that a forming process can be interrupted repeatedly without affecting the manner of its subsequent completion (Figures 9 and 10).

- (e) Negative creep can still take place at low temperatures (e.g. -183°C).
- (f) Negative creep is not noticeably affected by the presence of considerable amounts of iodine and mercury (added before melting) in the selenium.
- (g) The forming process is more effective for counter-electrodes of Sn-Cd alloy than for those of gold. Counter-electrodes of gold also have the lower barrier resistance.
- (h) The self-capacitance of the barrier layer decreases during the forming process.
- (i) A test on the basis of Schottky's capacitance method (as far as it is applicable) does not reveal large gradients in the concentration of impurity centres near the counter-electrode before or after forming.

Selenium is known to be a deficit semiconductor from the polarity of its thermoelectric power and from the direction of rectification. Its impurity centres thus have an electro-negative character. They are presumably attracted towards the counter-electrode by the normal diffusion field, and more strongly attracted when a reverse voltage is applied to the barrier layer. Since the field strengths involved are large, it is to be expected that these impurity centres will, in the course of time, be displaced towards the counter-electrode until a new equilibrium is established. In absence of an external field the same process should occur, though more slowly and not to the same stage of completion, i.e. the junction is expected to undergo a partial self-forming process before any reverse voltage is applied. Junctions which have counter-electrodes of apparently equal area are frequently found to have different resistances (per unit area) although prepared by identical methods. A self-forming process of the type here envisaged would account for such differences.

The diffusion field is always strongest in the immediate proximity of the counter-electrode and this is the region from which impurity centres are most likely to be displaced. The exact nature of the centres is still unknown. Since iodine, which is often added to selenium in the course of rectifier production, does not materially affect the creep process, it must be concluded that it does not form active impurity centres. Mercury (which is in any case more likely to be an electron donor than an acceptor) likewise does not appear to form such centres. These conclusions are supported by thermoelectric measurements on selenium which will be described in another paper.

What function can be ascribed to the displaced impurity centres (whatever their exact nature may eventually be found to be)? It is possible to envisage that these centres may separate out at the interface between the selenium and the counter-electrode, or may combine with the counter-electrode to form a stable compound provided the electrode is made of a suitable material. A chemical barrier layer would result in both cases. The normally good stability of fully formed rectifiers lends probability to the second hypothesis. A counter-electrode of Sn-Cd alloy would then behave rather differently from one of gold which is chemically almost inert. Such differences of behaviour have already been described. The gold counter-electrodes are associated with lower barrier resistances and a less effective forming process, though this may be partly due to differences in the effective contact areas. The two mechanisms for forming here suggested are not mutually exclusive. The separation of impurity centres may, and probably does, precede the formation of a compound with the counter-

electrode. Figure 16 shows schematically the various stages of the forming process as envisaged by the present theory.

Once an impurity ion has entered into combination with the counter-electrode, it is firmly held, but impurities which have merely separated out and have not yet entered into combination may, under suitable conditions, diffuse back into the selenium. Conditions would be particularly favourable for this when a forward voltage is applied. This would account for the observed de-forming produced by the prolonged passage of forward current through the junction. In the absence of a forward voltage the same process would take place, though again more slowly. This is probably the mechanism of what is commercially called 'ageing'.

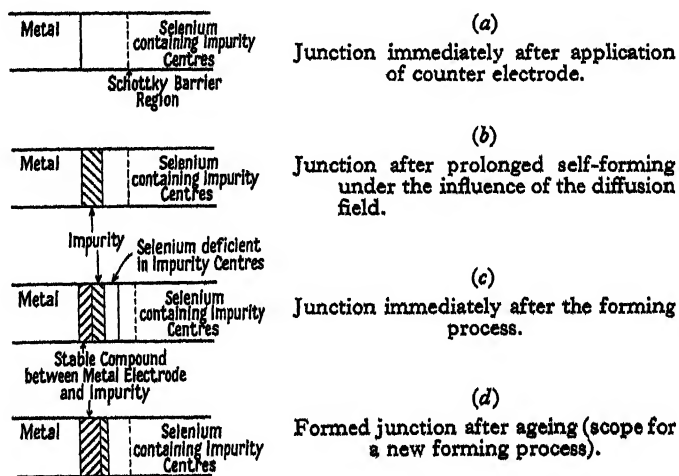


Figure 16. Schematic representation of the forming process as suggested by current creep measurements.

It is now necessary to consider how the forming process ends. On the basis of the present model, negative creep would be expected to cease eventually for two reasons: (a) because the field strength within the selenium decreases as the chemical barrier develops, and (b) because the region in the selenium nearest to the counter-electrode becomes exhausted of impurity centres. The experimental results suggest that the second limitation is the more important. During prolonged storage, particularly at higher temperatures, the supply of impurity centres could be replenished by diffusion from the selenium outside the barrier region and perhaps also by re-diffusion of the impurity deposit at the interface (see Figure 16(c)). If a high reverse voltage is then re-applied, a further forming process could take place, as shown in Figure 11.

One of the unexpected results is the observation that current creep can still take place at very low temperatures. Ionic motion in a crystal lattice is normally characterized by high activation energies and is therefore expected to be very temperature dependent. The present observations suggest an electrolytic process of remarkably low activation energy, as could only be envisaged if the selenium were in a low state of order, i.e. in a form intermediate between amorphous and crystalline. X-ray measurements tend to corroborate this interpretation.

ACKNOWLEDGMENTS

The authors would like to make grateful acknowledgment to Prof. R. W. Ditchburn for placing research facilities at their disposal and for much valuable help and advice. Thanks are also due to Mr. T. R. Scott of Standard Telecommunication Laboratories Ltd. for his interest in this work and for kindly supplying the test specimens.

REFERENCES

- HENISCH, H. K., 1949, *Metal Rectifiers* (Oxford: Clarendon Press).
 KOBAYASHI, A., 1949, *J. Phys. Soc., Japan*, 3, 41.
 ROSE, F., and SCHMIDT, H., 1947, *Z. Naturforschung*, 2a, 226.
 WILLIAMS, A. L., and THOMPSON, L. E., 1941, *J. Instn. Elect. Engrs.*, Pt. I, 88, 353.
 U.S.A. Signal Corps Engineering Laboratories, 1949, *Technical Memorandum M-1173*, Fort Monmouth, N.J., U.S.A., 31st January 1949.

Secondary Electron Emission from Molybdenum produced by Helium, Neon, Argon and Hydrogen

By D. GREENE

Department of Experimental Physics, Queen's University, Belfast

Communicated by K. G. Emeléus; MS. received 13th March 1950

To be read on 24th November, 1950

ABSTRACT. Experiments have been carried out on the secondary electron emission produced at very low pressure from a degassed metal surface by neutral particles resulting from the neutralization of positive ions of about 1,000 electron volts energy in a metal canal. The gases used were helium, neon, argon and hydrogen. For the first three the energy distribution of the secondary electrons is compatible with the view that a large part of the emission is due to communication of the energy of excitation of a metastable atom to a metallic electron. With hydrogen there is evidence that a similar process occurs with normally excited states of the atom or molecule. In all cases there is evidence that some emission occurs through local heating or disruption of the metal lattice by the mechanical impact of swift atoms, the extent of this increasing from helium and neon to argon and hydrogen. Some reflection of the swift atoms occurs from the metal target.

INTRODUCTION

THE secondary electron emission produced by metastable helium atoms at a degassed metal surface in a vacuum has been investigated by Oliphant (1929). He found that the energy distribution of the electrons had an upper limit, equal to the energy of excitation of the metastable atom, less the work function of the metal, and a lower limit, which agreed well with the value of the energy of excitation of the metastable atom less a calculated value for the total potential barrier at the surface of the metal. These observations seemed to show that the secondary electron emission could be regarded as the result of individual interactions between the incident metastable atoms, and the electrons in the metal. The theory of the energy distribution amongst the secondary electrons ejected in this way has been developed by Massey (1930, 1931). In Massey's theory the metal is treated as a molecule whose ionization potential is equal to the work function, and the interaction considered as a collision of the second kind. The theoretical energy distribution shows a higher proportion of high energy electrons than was found by Oliphant, and a sharper cut-off at the upper energy limit. Massey expressed the opinion, however, that the agreement

with experiment was as good as could be expected from the approximate nature of the calculations and the difficulties of the experimental work.

Chaudhri and Khan (1948) have studied the secondary emission produced by neutral atoms or molecules of mercury and potassium. Although mercury has metastable states, their excitation energy is little greater than the work function of the target, so that the source of energy for individual interactions of the kind considered by Massey is small. Potassium has no metastable state and therefore no internal energy available for such interactions, under the conditions of experiment. Chaudhri and Khan found that the upper electron energy limit for both potassium and mercury increased as the kinetic energy of the incident particles was increased, and that a semi-logarithmic plot of current of secondary electrons to a collector against retarding voltage gave nearly a straight line over a range of several volts. They interpreted this as showing that the emission was caused by highly localized heating of the target, due to transfer of the kinetic energy of the particles of the incident beam. From the slope of the line obtained by plotting the logarithms of the collector current against the collector potential a temperature was deduced for the secondary electrons, which could be taken to be the same as the temperature of the heated spot from which the thermionic emission was occurring. A similar explanation had been given previously of the secondary electron emission produced by fast alkali ions (Oliphant and Moon 1930).

It thus seems that the secondary electron emission may be due to not less than two processes: (i) individual interactions which are similar to collisions of the second kind; (ii) interactions where the kinetic energy of the incident particles causes a localized excitation or disruption of the metallic lattice, resulting in what can be described formally as a thermionic emission. The disruption of the lattice could be due either to the impact of a single particle or to successive impacts. In the former case the electron temperature would be independent of the intensity of the primary beam, while in the latter one would expect the electron temperature to increase with intensity. There is no experimental evidence available on this point but it is thought that the single impact process is the more likely. In this connection it should be noted that Oliphant (1929) observed that sputtering was produced by the beam of metastable helium atoms, showing that some disruption of the lattice did occur; even in this later experiment therefore some of the electron emission observed may thus have been thermionic.

The main purpose of the present experiments was to confirm Oliphant's results for helium, and to extend them to neon and argon. The results which have been obtained are in general agreement with Oliphant's, but more evidence has been found that emission due to localized heating takes place. Some experiments have also been carried out with hydrogen, which have given results in agreement with those of Chaudhri and Khan, and in addition show that electron emission may be produced by impact of atoms or molecules in ordinary excited states.

The experiments of Oliphant, Chaudhri and Khan, and the present author are not directly comparable with others, of which those of Uyterhoeven and Harrington (1930) are typical, where the target surface was located in a discharge plasma, since secondary electron emission is very sensitive to gas contamination of the surface.

§ 1. APPARATUS AND EXPERIMENTAL PROCEDURE

The final form of the experimental tube is shown in Figure 1. Gas was admitted to the bulb O, where an arc discharge of a few tenths of an ampere was passed between the spiral filament F_1 and the anode A. The electrode P was made strongly negative to the anode so that positive ions were drawn to it, some passing down a metal canal 5 cm. long and 2 mm. in diameter, into the region H, which was maintained at a hard vacuum by a high-speed oil diffusion pump of the type described by Sloan, Thornton and Jenkins (1935). Oil molecules were prevented from diffusing back into H by water-cooled baffles in the throat of the pump and a recessed liquid oxygen trap in the pumping lead. The field, applied between A and P, was reversed between the two coaxial cylinders P_1 and E_1 so that positive ions emerging from the canal were prevented from passing down the tube. Secondary electrons or negative ions produced in the canal would be accelerated to E_1 and, to remove these, the same potential

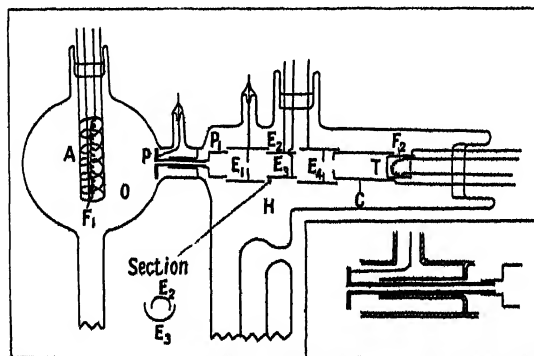


Figure 1. Experimental tube. Inset: details of canal on larger scale.

difference as was applied between P_1 and E_1 was applied between the electrodes E_2 and E_3 , two smaller semi-cylinders. A, E_1 and E_3 were at the same potential. The target T was of molybdenum, and formed the lid of a pill box which contained the filament F_2 . This was insulated from the pill box and could be used to raise the target temperature to bright red heat by radiation and electron bombardment. The cylinder C was used to collect charged particles from the target while E_4 , a cylinder containing a central diaphragm, served as a shield for C. E_4 and C were also at the same potential as E_1 and E_3 . The electrodes were all made of nickel except the canal P which was aluminium, the tungsten filaments and tungsten leads through the Pyrex glass.

The potential applied to accelerate positive ions from the arc plasma into the canal was provided by a battery of accumulators of up to 1,500 volts connected between the anode A and the canal P through an 80,000-ohm ballast resistance in the lead to the canal. The applied voltage was measured with an electrostatic voltmeter. All other potential differences were also provided by lead accumulators.

The secondary electron current between target and collector was measured with a galvanometer of sensitivity 5,000 mm/ μ a. at 1 metre. The total current entering the target-collector system was measured by a galvanometer approximately one-tenth as sensitive. The latter showed zero current during all the secondary electron experiments.

The helium, neon and argon used were obtained from the British Oxygen Company as spectroscopically pure. The hydrogen used was admitted through a palladium leak.

The gas pumped from the high vacuum side of the tube could either be circulated back to O, or pumped out into the air by a mercury diffusion pump backed by a 'Metrovac' mechanical pump. Traps cooled with liquid oxygen were inserted in all the leads to the experimental tube.

Before admitting gas to the system it was evacuated to a pressure of 10^{-6} mm. Hg or less by the auxiliary mercury diffusion pump, and the glass work was torched. The pressure was measured with a McLeod gauge which could be connected to either side of the experimental tube. The target was degassed by maintaining it at bright red heat in the best attainable vacuum. It was found that under these circumstances the thermionic current from the target consisted mainly of positive ions. This emission decreased by a factor of several hundred over a period of about seven hours, after which it reached a steady value of the order of 10^{-8} amp/cm². Subsequent reheating of the target, even after it had been exposed to the atmosphere, brought it back to the steady state in about half-an-hour. Before each set of observations on the secondary electron emission the target was degassed in this way for at least this length of time. The target temperature was then reduced to the maximum at which thermionic emission was negligible; this corresponded to a state of dull red heat.

The neon, argon and hydrogen were allowed to stream through the apparatus. The helium was circulated and, to prevent the accumulation of impurities, was continuously purified by passing it through activated charcoal cooled in liquid oxygen before it entered the arc. To avoid contamination of the target, the gas pressure in the target side of the tube must be kept as low as possible. The minimum usable pressure was determined by the pressure required in the arc to give a sufficiently high ion concentration. This was found to vary with the different gases. The upper usable pressure on the high vacuum side of P is the pressure required to give a mean free path equal to the path distance of the neutral particles. These two factors restrict the range of working pressures in O. For neon and argon the main observations were made with a pressure of 8×10^{-5} mm. Hg in the target side of the tube, while for helium and hydrogen the working pressures were 6×10^{-4} mm. Hg and 2×10^{-5} mm. Hg respectively. The corresponding gas pressures in the arc were 10^{-1} – 10^{-2} mm. Hg.

In addition to the tube described several others were tried. The details of the electrode structures were different, but the main results obtained were the same in all cases. One of the earlier tubes had a longer neck (5 cm.) than that shown in Figure 1 between the bulb O and the region H. Some results obtained with the long-necked tube will be referred to in §2.

§2. PRELIMINARY OBSERVATIONS

In a preliminary experiment, the electrodes E_1 , E_2 , E_3 and E_4 were joined to P so that positive ions could reach T and C, which were joined together. The energy distribution of the positive ions was investigated by the application of a variable retarding potential difference between T and E_4 . The positive ion current was found to cut off within 3 volts when the retarding voltage approached the accelerating voltage, and a small reverse current which saturated rapidly was observed when the retarding voltage was further increased. This showed (i) that positive ions could not reach the target under the conditions of the main

experiment; (ii) that the positive ions coming from the canal had an energy spread of between two and three volts. The small reverse current was presumably due to secondary electrons produced by positive ions colliding with the edges of the slits in the electrodes E_1 and E_4 . With the field between E_2 and E_3 as in the main experiments, these secondary electrons could not reach the target. The small energy spread of the positive ions indicates that few gas collisions are taking place in and beyond the canal, and that the particles producing the secondary emission in the main experiments came direct from the arc or resulted from collision between positive ions and the walls of the canal. From the geometry of the tube it is likely that these particles are neutral atoms formed by positive ions capturing electrons from the metal of the canal wall, in making glancing collisions with it. Further evidence for this was found using the long-necked experimental tube (§ 1). On applying a localized magnetic field perpendicular to the axis of the canal, the total secondary electron current from the target was found to vary with the magnetic field strength, when only neutral particles could reach T, in the manner shown in Figure 2. The magnetic field had no effect on the measured arc current or on the appearance of the luminous plasma in O. The explanation of this curve seems to be as follows. As the magnetic field is increased from zero the positive ions passing down the canal

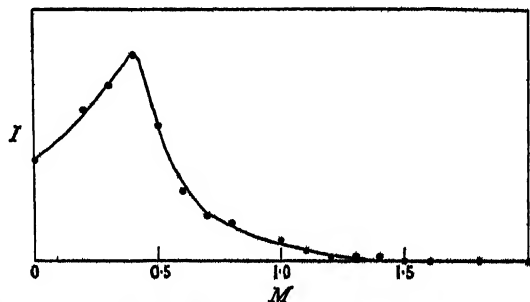


Figure 2. Effect of magnetic field perpendicular to canal; M = magnet current; I = total secondary electron current, arbitrary units. Magnetic field for peak current, approx. 90 gauss.

have their paths increasingly bent so that the number of collisions with the walls is increased, producing an increase in the number of swift neutral particles emerging into the high vacuum. Further increase in the field bends the paths of the positive ions so much that their collisions with the walls are no longer glancing. This results in a decrease in the number of neutral particles able to emerge and therefore in the secondary electron current from the target. The fact that the decrease to zero of the secondary current is not sharp may be due to the non-uniformity of the magnetic field, which was necessary in order to localize it. The curve was reproducible in its main features. It was not possible to obtain it for the final tube, in which the short neck made it impossible to apply the magnetic field without disturbing the arc.

These results support Oliphant's conclusions that the particles producing the secondary electron emission are metastable atoms formed by glancing collisions of positive ions with the canal walls, and that the secondary emission is due to neutral atoms rather than to quanta coming from the arc. It will, therefore, be assumed that in the present experiments emission due to quanta was negligible for all the gases used, as the experimental conditions were closely similar to those in these preliminary experiments with argon, and in Oliphant's experiments. It is also reasonable to assume that the kinetic energies of the

neutral atoms formed by glancing collisions will have values up to that of the incident ions. Further, when the experimental tubes were dismantled, each target was found to have a well defined spot where it had been struck by the beam. In the first tube where the target was not adequately degassed this spot was discoloured, while in later tubes the spot was not discoloured but was pitted. This could hardly have been caused by photons.

In all the experiments a slight instability in the secondary electron current from T to C was observed. This corresponded to fluctuations of the order of 10^{-10} amp. It showed that it would have been pointless to aim at higher sensitivity, which would otherwise have been desirable. It is thought that these fluctuations may have been connected with the fact that it was not possible to degas the inner surface of the canal. To maintain the differential pumping the probe had to be held in a tight fitting glass sheath and therefore could not be treated by the usual methods. The positive ions colliding with the canal walls would cause a localized degassing, but under the circumstances this may well have led to variations in conditions at the canal surface which would produce fluctuations in the intensity of the neutral beam.

§ 3. MAIN OBSERVATIONS

(a) *Inert Gases*

The procedure in the main experiments was the same as in those of Oliphant, namely to observe the variation in the electron current between target T and collector C, when the potential difference between them was altered. Retarding

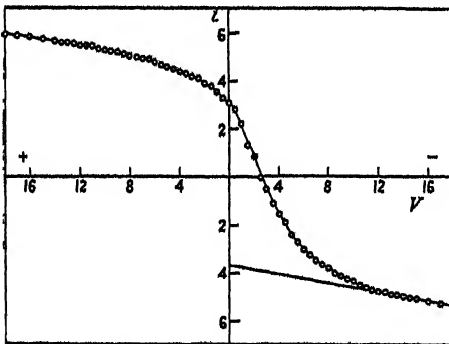


Figure 3. Retarding curve for secondary electron current given by neon atoms. V = potential of collector C with respect to target T; i = secondary electron current, unit 2×10^{-8} amp.

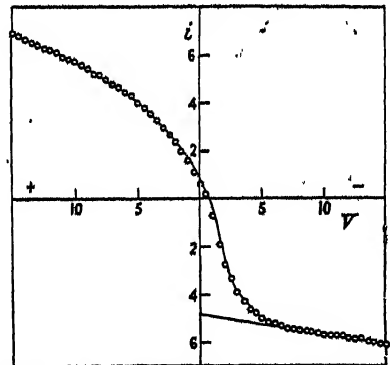


Figure 4. Retarding curve for secondary electron current produced by argon atoms. Unit current = 2×10^{-8} amp.

curves were obtained in this way for different values of the kinetic energy of the primary beam.

By comparing the results it was hoped to obtain information on the relative importance of the processes (i) and (ii) discussed in the Introduction. It was found that as the energy of the incident particles was lowered (i.e. as the potential difference applied between A and P (Figure 1) was decreased) the total secondary electron current from T to C decreased. This and the limitation on the sensitivity already discussed (§ 2) set a lower limit to the energy of the beams at which measurements could be made.

The results for typical measurements made with neon, argon and helium are shown in Figures 3, 4 and 5. These are similar to those obtained by

Oliphant (1929) for helium. The fact that the current between the target and collecting cylinder changes sign when the potential difference between them is reversed was interpreted by him as showing that some of the incident metastable atoms are reflected from the target, and cause secondary emission from the surface of the collecting cylinder. The differential curve obtained from the right-hand side of the curve gives the energy distribution of the secondary electrons from the target. Figure 5 for helium shows only this part of the curve. It will be seen that the secondary electron current from the target does not saturate, but becomes a straight line of small constant slope when the voltage is sufficiently increased. This may be due to (1) the unsymmetrical structure of the target-collector system; (2) a slight steady change in the gas pressure in the arc during the time over which a set of observations were being made. The slope of the straight line was different in different runs, and to make the results

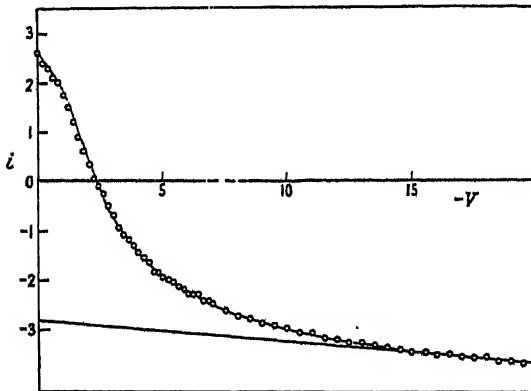


Figure 5. Part of retarding curve for electron current produced by helium atoms. Unit current = 2×10^{-9} amp.

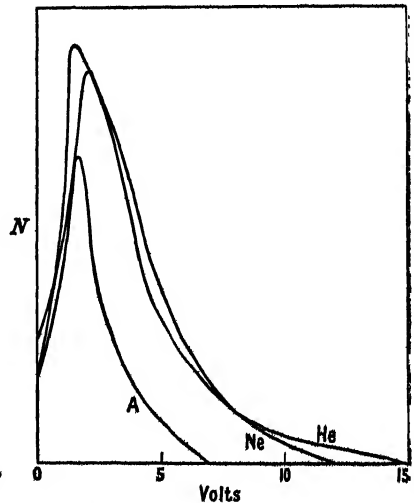


Figure 6. Electron energy distribution curves corresponding to Figures 3, 4, 5; E = electron energy; N = no. of electrons per unit energy range in arbitrary units.

comparable, the straight line was extrapolated back to zero volts and the retarded electron current taken to be the difference between this extrapolated line and the experimental curve. Because of the unsymmetrical nature of the target-collector system the left-hand side of the curves, which corresponds to an electron current from C to T, does not give much information. Further, some of the incident particles presumably lose kinetic energy on being reflected, and this may also affect the shape of the left-hand side. The lack of symmetry of the target-collector system does, on the other hand, tend to eliminate effects due to multiple reflections between the target and cylinder, so that the ratio of the saturation currents at the beginning of the straight sections is a rough measure of the fraction of the incident metastable atoms reflected from the target. These retarding curves were found to be closely reproducible when the target was degassed by the process described (§ 1). The electron energy distribution curves derived from the retarding curves in Figures 3, 4 and 5 are shown in Figure 6.

To make the retarding curves obtained using incident particles of various kinetic energies comparable, it is desirable to adjust them so that they have the same maximum ordinate. Families of observations for various energies, adjusted in this way, are shown in Figures 7, 8 and 9.

The energy range which could be covered for helium metastable atoms was rather small, viz. 730 to 1,080 electron volts, chiefly because of the small secondary emission. Within this range the form of the retarding curve was found to be independent of the kinetic energy of the incident particles, as shown by Figure 7.

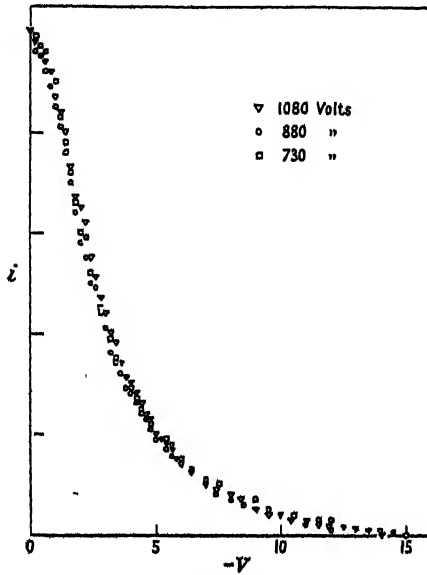


Figure 7. Adjusted secondary electron currents produced by helium.

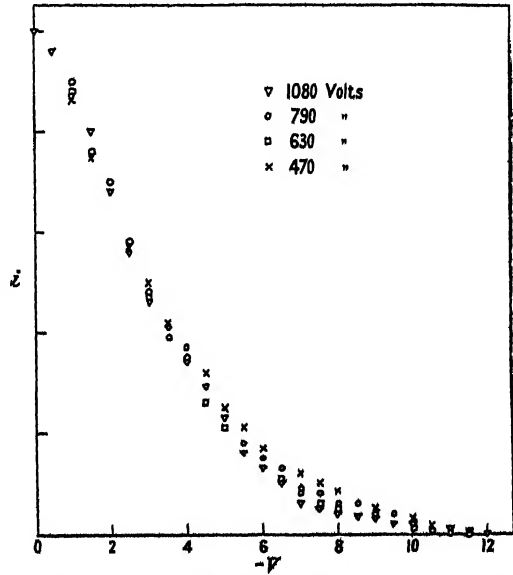


Figure 8. Adjusted secondary electron currents produced by neon.

Figure 8 shows some results for neon, taken over the energy range 470–1,080 electron volts. Here the spread of the experimental points is slightly wider than in Figure 7, but the random placing of the points with respect to the energy of the primary particles seems to show that, within experimental error, the form of the curve is almost independent of this energy.

Figure 9 shows typical results for argon. Here in the range of 0 to 4 collector volts the spread of the experimental points is quite wide, and the points lie in the order of the primary kinetic energies to which they correspond. Above -4 volts the points converge towards a common limit, and their placing becomes more nearly random. Two sets of observations made with incident particles of the maximum and minimum energies used are shown. In both sets of results the differences between corresponding points is smaller than the difference between the corresponding points taken at the next value of the kinetic energy of the primary particles. This shows that the spreading of the experimental points taken at different kinetic energies is outside experimental error.

(b) Hydrogen

A typical retarding curve obtained using hydrogen is shown in Figure 10. This shows the same general features as were obtained with the inert gases. The corresponding velocity distribution curve is superimposed. A family of retarding

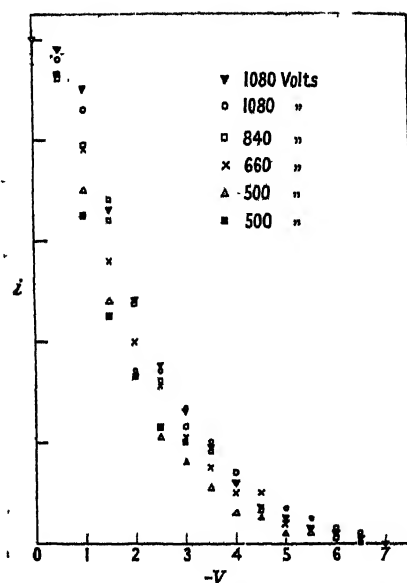


Figure 9. Adjusted secondary electron currents produced by argon for different values of kinetic energy of the primary beam. V =retarding voltage; i =secondary electron current from T.

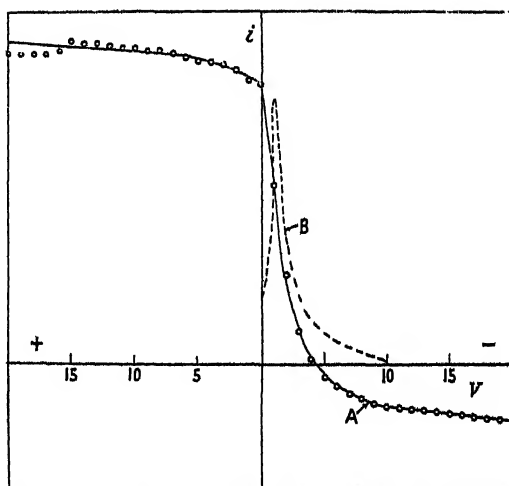


Figure 10. Typical retarding curve A, and derived electron energy distribution curve B for hydrogen.

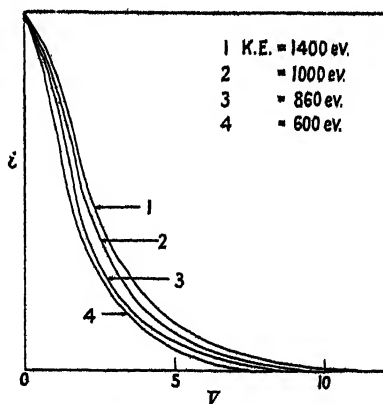


Figure 11 Adjusted secondary electron currents produced by hydrogen atoms and molecules of various kinetic energy.

curves for hydrogen, adjusted to a common maximum ordinate, is shown in Figure 11. The experimental points lie closely on the curves. The upper limit of the energy of the secondary electrons increases with the kinetic energy of the primary particles, and the experimental points follow the order of the primary particle energy, for all values of the collecting field.

§ 4. DISCUSSION OF RESULTS

(a) *Inert Gases*

The maximum energy a metastable atom can give to an electron in the metal, apart from any contribution due to its kinetic energy, is given by the energy of the metastable state less the work function of the metal. These values for helium,

neon and argon are compared with the observed upper energy limits of the secondary electrons in Table 1. The accuracy of the observed values for the upper energy limits is limited to the nearest 0.5 volt by two factors: (i) the difference between successive galvanometer readings at 0.5 volt intervals at this part of the curve is approaching the least measurable change in current; (ii) it is difficult to decide exactly at what point on the curve saturation is reached, because the approach to saturation seems to include a small asymptotic contribution. With these limitations, the calculated and observed upper energy limits are in good agreement.

There is no definite evidence for secondary emission from atoms in the 20.5-volt metastable state of helium, but the resolution was probably insufficient to show any change in slope of the experimental line arising in this way.

The lower limit of the secondary electron energy should be given (Oliphant 1929) by the energy of the metastable state less the total potential barrier at the surface of the metal. The latter has a calculated value of 17.9 electron volts (Oliphant 1929); this would give a minimum electron energy of zero for neon and argon, and is in agreement with the observations. On the other hand, the

Table 1

Gas	He	Ne	Ar
Energy of metastable state (ev.)	19.8	16.5	11.5
Work function (v.)	4.4	4.4	4.4
Excess energy available (ev.)	15.4	12.1	7.1
Observed upper energy limit (ev.)	15.0	12.0	7.0

minimum value predicted for the electrons liberated by helium is 1.9 electron volts, while the observed minimum was zero. There may be several explanations for this: (i) the potential barrier at the surface of the metal may not have the value assumed; (ii) if the electrons were scattered in emerging from the surface, this might reduce their energy; (iii) secondary electrons liberated by the kinetic energy of the incident particles would have zero minimum energy, and the presence of these might mask the lower energy limit for individual interactions; (iv) a contact potential between the target and collecting cylinder might shift the voltage scale enough to mask the presence of a non-zero minimum energy; it is unlikely that this could account completely for the failure to observe the predicted minimum energy, because the presence of a contact potential difference of the magnitude required would also produce an appreciable alteration in the value of the observed upper energy limit.

The consistency of the results obtained using helium metastable atoms of various kinetic energies would seem to show that the kinetic energy of the particles made no appreciable contribution to the observed secondary electron emission. The form of the observed energy distribution curve is nearer to that obtained by Oliphant than to that calculated by Massey. The results for neon are also almost independent of the kinetic energy of the incident metastable atoms. The results for argon show a definite tendency for the proportion of the higher energy electrons in the region 0-4 electron volts to increase with increase in kinetic energy of the bombarding particles. For still higher energies of the secondary electrons, the energy distribution curves converge to a common upper limit. It therefore seems likely that the kinetic energy of the argon metastable atoms does contribute to the observed secondary emission by the process (ii) discussed

in the Introduction. We would expect the mean energy of the electrons liberated in this way to increase with increase in the kinetic energy of the incident particles, and this also is in agreement with the observed results. Apparently these 'thermionic' electrons have not enough energy to change appreciably the high energy tail of the energy distribution. The difference between the results for argon and those for neon and helium could be due either to the greater mass of the argon atom or to the smaller excess energy available for individual interactions between argon metastable atoms and metallic electrons.

The ratio of the saturation current from C to the sum of the negative and positive saturation currents from C and T gives an approximate measure (§ 2) of the fraction of the metastable atoms reflected from the target. Its magnitude for different values of the primary kinetic energy is shown in Figure 12. For all

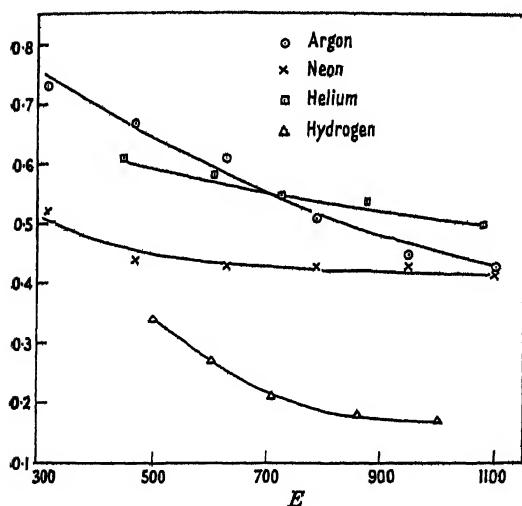


Figure 12. Fraction of incident particles reflected from target as a function of incident kinetic energy, E volts.

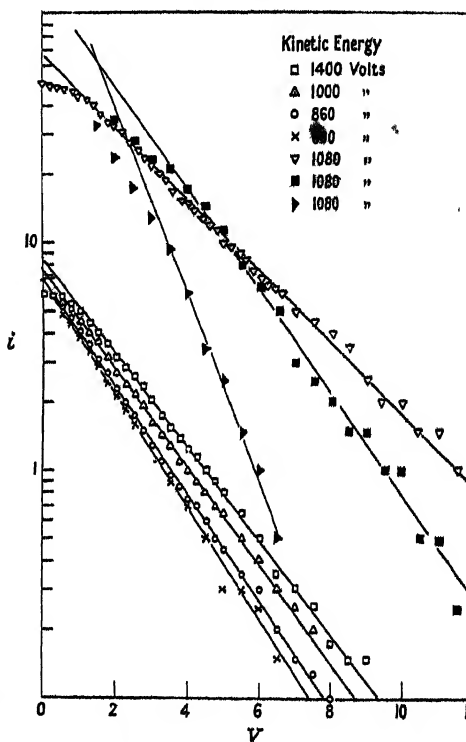


Figure 13. Semi-logarithmic plots of retarded electron currents i against retarding voltage V . The top four symbols refer to hydrogen, the remainder to helium, neon and argon in descending order.

the gases the proportion of the incident particles reflected from the target decreases with increase in their kinetic energy. This effect is most marked for argon, and may be connected with the greater variation in the form of the secondary electron energy distribution.

(b) Hydrogen

The beam of neutral particles formed in the present apparatus almost certainly contained both atoms and molecules, and investigations concerned with the design of proton sources (Lamar, Samson and Compton 1935) show

that it might be predominantly molecular. The results obtained with hydrogen are different from those given by the inert gases, in that the whole of the retarding curve shifts with the kinetic energy of the incident particles. Semi-logarithmic plots of the retarding curves were found to give good straight lines. Some of these are shown in Figure 13. For comparison semi-logarithmic plots of some retarding curves taken with the inert gases are also shown. Those for argon and neon show a greater departure from linearity than those for hydrogen and helium.

The fact that the semi-logarithmic plots for hydrogen are linear over nearly the complete range of voltages may not be very significant because any rapidly falling retarding curve would tend to give such a result. Assuming, however, that the electron emission is of the second type discussed in the Introduction the electron temperature can be determined from the slope of the semi-logarithmic plot, and this in turn can be taken to be the effective temperature of the emitting surfaces. This description of the results is only an approximation as an energy distribution with a finite upper limit is not truly Maxwellian.

The results obtained for hydrogen are summarized in Table 2. These show that the electron temperature increases with the energy of the incident particles, though not very uniformly, and seems to be approaching a limiting value. The upper energy limits also seem to approach a limiting value of about 11 or 12 volts.

Table 2

Max. energy of incident particles (ev.)	Electron temp. (° K.)	Mean electron energy (ev.)	Mean electron temp. (° K.)	Max. electron energy (ev.)
600	20,600	2.7	20,600	7.5
710	21,500	2.8	21,500	8
860	21,500	2.8	21,500	8.5
1000	24,800	3.2	25,100	9
1020	23,800	3.1		10
1020	25,400	3.3		10
1000	26,700	3.4	25,000	10
1150	25,000	3.25		11
1350	24,000	3.1		12
1400	27,800	3.6	25,100	11
1400	24,700	3.2		11

It can be seen from Figure 13 that all the semi-logarithmic plots depart from linearity in the neighbourhood of zero collecting volts, and that this effect becomes more pronounced as the energy of the incident particles is increased. Although the zero point in these curves, as for the inert gases, is a little indefinite, we would not expect this to vary with the kinetic energy of the primary particles, where the total power delivered by the beam is small; it therefore seems likely that the effect is real. This suggests that more than one process is involved in the production of the secondary electrons.

So far as the author knows, the possibility that direct interactions between ordinary excited atoms and metallic electrons could produce secondary emission in the present type of tube has not previously been considered. If the time required for the primary particles to travel from the canal (where they are neutralized) to the target were of the same order of magnitude as the mean life-time of an ordinary excited state, and if excited atoms or molecules were produced by neutralization, such interactions would be possible and could cause secondary

electron emission if the energy of excitation were greater than the work function of the metal. Calculated values of the mean life-times of the excited states of hydrogen atoms are given by Condon and Shortley (1935). The mean life-times increase with the degree of excitation, the value given for $n=6$, for example, being 1.96×10^{-7} second. This value is obtained by assuming that equal numbers of atoms exist in all the states corresponding to $n=6$. The life-times of the individual states may be considerably greater than this value, for example, the value given for the 6h state is 6.1×10^{-7} second. The transit times for 1,000 and 1,400 electron volt hydrogen atoms travelling between the canal and target (a distance of 10 cm.) are 2.88×10^{-7} and 1.92×10^{-7} second respectively. It would, therefore, be possible for some hydrogen atoms in the higher excited states to reach the target before decay. The excitation potentials of these states go up to 13.54 volts, the ionization potential, and since this is well above the work function of molybdenum it seems that they could produce secondary emission by interactions of type (i). Photons produced by the decay of excited atoms in transit between the canal and surface could also produce electron emission. Only a small proportion of these photons would, however, reach the target. Secondary electrons produced by direct interactions with excited atoms might thus be responsible for the non-Maxwellian contribution to the total emission for the greatest secondary electron energies.

The maximum energy an excited hydrogen atom could give a secondary electron is the ionization energy less the work function of the target, i.e. 9.1 electron volts. This could not account for the upper energy limit of 11–12 electron volts to which the results tend. The ionization potential of the hydrogen molecule is 15.6 volts, so that the maximum energy an excited hydrogen molecule could give to a metallic electron is 11.2 electron volts. This agrees well with the observed result. The mean life-times of the excited states of the hydrogen molecule are not known, but if, as is likely, they are of the same order of magnitude as those of the atom, secondary emission by individual interactions would be possible.

Since the transit times for helium atoms are twice those for hydrogen atoms, it seems possible that the ordinary excited states might cause secondary emission for the higher kinetic energies of the bombarding particles. There is, however, no definite evidence for this from the present work. This might be due to a tendency of the neutralization process to leave the helium atoms in the metastable state. Neon and argon would probably give transit times which are too long to allow any effects due to excited atoms.

A further mechanism that has to be considered for the ejection of secondary electrons is that in which elastic collisions between incident atoms and metallic electrons give the latter sufficient energy to penetrate the potential barrier after reversing their direction of motion in an elastic collision. The maximum energy a hydrogen atom with kinetic energy of 1,000 electron volts could give an electron in an elastic collision is about two electron volts; since this is less than the work function of the metal, it would seem that this mechanism can be ruled out.

The retarding runs taken with hydrogen give similar evidence to that obtained with the inert gases for reflection of the incident particles from the target. The reflection coefficients determined as before are shown in Figure 12. The existence of an electron current from the collector to the target in this case might also be due in part to photons produced by excited atoms in the beam decaying inside the collecting cylinder.

§ 5. SUMMARY

The results reported here generally confirm those of Oliphant and of Chaudhri and Khan. The electron energy distribution given by helium metastable atoms agrees with that observed by Oliphant rather than that calculated by Massey. The present results for hydrogen are comparable with those of Chaudhri and Khan for potassium, because neither atom has a metastable state, and both agree in showing the secondary electrons to have a nearly Maxwellian energy distribution. The results for neon are similar to those for helium in that the form of the electron energy distribution does not vary appreciably with the energy of the primary particles, while the proportion of high energy electrons liberated by argon metastable atoms increases with their kinetic energy. The similarity of the results for hydrogen to those of Chaudhri and Khan (1948) for mercury is of particular interest, because while it is known that metastable atoms of thermal energy can liberate secondary electrons (Webb 1924), the emission produced by high-speed mercury metastable atoms seems to be due almost entirely to their kinetic energy.

The general conclusion that can be drawn from the data at present available seems to be that the relative efficiency of the individual interaction process, and the mechanical transfer process, varies with the excess energy available for the former. For neon and helium the contribution made by the second process is negligible; for argon there is a small proportion of the secondary emission caused by the second process; for mercury, where the excess potential energy available has a slight positive value the emission seems to be almost entirely due to the quasi-mechanical transfer process, while hydrogen and potassium give results which are also explicable in terms of the latter process. There is also evidence that hydrogen atoms and molecules can produce secondary electrons by individual interactions, where ordinary excited atoms can reach the target. These conclusions again are in agreement with work on the secondary electron emission produced by positive ions (Oliphant and Moon 1930), where the relative importance of the emission produced by individual interactions and by the kinetic energy of the incident particles varies in the same way.

ACKNOWLEDGMENTS

The author wishes to thank Professor K. G. Emeléus for suggesting and supervising this work, Dr. R. H. Sloane for advice on many points, and Professor H. S. W. Massey for a stimulating discussion. The work was done whilst the author was a Senior Research Student of Queen's University. Part of the cost of batteries was defrayed from the Research Fund of the University.

REFERENCES

- CHAUDHRI, R. M., and KHAN, A. W., 1948, *Proc. Phys. Soc.*, **61**, 526.
CONDON, E. U., and SHORTLEY, G. H., 1935, *The Theory of Atomic Spectra* (Cambridge: University Press), p. 346.
LAMAR, E. S., SAMSON, E. W., and COMPTON, K. T., 1935, *Phys. Rev.*, **48**, 886.
MASSEY, H. S. W., 1930 *Proc. Camb. Phil. Soc.*, **26**, 386; 1931, *Ibid.*, **27**, 460.
OLIPHANT, M. L. E., 1929, *Proc. Roy. Soc. A*, **124**, 228.
OLIPHANT, M. L. E., and MOON, P. B., 1930, *Proc. Roy. Soc. A*, **127**, 388.
SLOAN, D. H., THORNTON, R. L., and JENKINS, F. A., 1935, *Rev. Sci. Instrum.*, **6**, 75.
UYTERHOEVEN, W., and HARRINGTON, M. C., 1930, *Phys. Rev.*, **35**, 438, and **36**, 709.
WEBB, H. W., 1924, *Phys. Rev.*, **24**, 113.

The Variation with Temperature of the Piezoelectric Coefficients of Quartz

By A. C. LYNCH

Post Office Engineering Research Station, London N.W.2

MS. received 14th June 1950

ABSTRACT. The equivalent electrical circuits of three bars in longitudinal vibration were measured at approximately 25, 52.5 and 80° c. In this range of temperature the temperature coefficients of d_{11} and d_{14} are respectively -2_{80} and $+13_{80}$ parts/million/° c., and there is no evidence to support Cady's suggestion that d_{11} passes through a maximum near room temperature.

The measurements suggest a rather high value for d_{14} : $(-2.21 \pm 0.1) \times 10^{-8}$ cm/E.S.U. of potential at 25° c.

§ 1. INTRODUCTION

WIDELY varying results have been reported for the piezoelectric coefficient d_{11} of quartz at high and low temperatures, while no values have been published for the variation with temperature of d_{14} (Cady 1946, p. 221). A few results, expressed as temperature coefficients, are :

Observers	Date	Range of temperatures (°c.)	Temperature coefficient of d_{11} (parts/million/°c.)
Andreeff, Fréedericksz, and Kazarnowsky	1929	15 to 500	-350
Van Dyke	1932	-80 to 40	Small at -80°c., -1000 at 40°c.
Langevin	1936	20 to 200	-550
Langevin and Moulin	1937	-60 to 23	Zero between 0 and 20°c., +1000 below 0°c.
Clay and Karper	1937	17 to 90	-110

Cady quotes the last of these as -7, owing to a misreading of Clay and Karper's notation. He concludes, from these and other reported results, that the coefficient probably passes through a very flat maximum near room temperature. Since there is no crystallographic transition to cause a discontinuity in the variation, we might expect a roughly parabolic curve for piezoelectric coefficient against temperature. If so, measurements at three temperatures should show whether a maximum exists, not only if it lies between the temperature of measurement, but also if it is somewhat outside them. Apparatus was available with which the equivalent electrical circuit of quartz resonators could be measured with sufficient accuracy in the temperature range 25 to 80°c.

In these measurements we need not attempt to find the absolute value of the piezoelectric coefficient, but only its variation with temperature. Hence we need not consider the approximations made in deducing it from the equivalent circuit, nor any mechanical effects introduced by the electrodes, and the specimens need not be those best suited to the measurement of the piezoelectric coefficient itself. If the measurements lead to a value for d_{11} agreeing with the accepted figure of 6.9×10^{-8} cm/E.S.U. of potential (Cady 1946, p. 219)—as in fact they did—then we know the approximations are sufficiently good at 20°c.; we are then assuming only that they are equally justified at other temperatures. If the

experiments had disclosed a non-linear variation of d with temperature, this assumption would need further examination; but as d was found to vary linearly, there is no reason to doubt the assumption.

§2. EXPERIMENTS

For the sake of simplicity in interpreting the results, three bars were chosen for test, one having its length along the conventional Y-axis, and the others at 45° and at 135° to the Y-axis in the YZ plane. If an alternating electric field is applied along the X-axis, the bars can be set into longitudinal vibration; the behaviour of the Y-axis bar then gives a measurement of the coefficient d_{12} , which in quartz is equal to $-d_{11}$, while the results for the other two bars, taken together, give values for both d_{11} and d_{14} .*

Table 1 gives their dimensions and frequencies of resonance.

Table 1

Direction of length of bar	Dimensions (cm.)	Approximate resonance frequency (kc/s.)
Y-axis	$3.622 \times 0.503_s \times 0.187_s$	75
45° to Y-axis	$2.001 \times 0.300_s \times 0.204_s$	139
135° to Y-axis	$2.001 \times 0.300_s \times 0.204_s$	174

Silver was evaporated on to the faces in the YZ plane to serve as electrodes and the bars were held between domed silver contacts mounted on light springs. Thus mounted, the bars were set up in a lagged copper box, heated by a current of warm air. Temperatures were measured by a mercury thermometer mounted with its bulb close to the specimen, and could be kept within about 0.2 deg. for periods of a quarter of an hour. The specimens took up each new temperature within a few minutes (as judged by observations of their frequencies of series resonance), but were left at approximately the required temperature for about two hours before measurements were begun.

Measurements of equivalent capacitance at frequencies near that of series resonance were made as described by Lynch (1950). The only difficulty was that the oblique bars had large temperature coefficients of frequency, and special care was needed to keep their temperatures constant and to work quickly. Closely reproducible results were eventually obtained, however, and are given in Table 2.

The capacitances have been corrected, by interpolation, to values for three uniformly spaced temperatures. C is the measured capacitance in the equivalent circuit in pF.; f_0 is the frequency of series resonance in kc/s.; d' is the piezoelectric coefficient deduced†, with positive or negative sign chosen to make d_{11} positive; d_{14} and the second value for d_{11} are obtained from d' for the oblique bars; d' , d_{11} , and d_{14} are in 10^{-8} cm/E.S.U. of potential. The density of quartz has been taken as 2.64_s at 20° C. (Cady 1946, p. 411), which is approximately

* Cady (1946, p. 212) gives, for a bar rotated through an angle θ round the X-axis, $d_{12}' = -c^2 d_{11} + s c d_{14}$ where $c, s = \cos \theta, \sin \theta$. Hence, for the two specimens mentioned, $d_{11}' = -[(d')_{45} + (d')_{135}]$ and $d_{14}' = (d')_{45} - (d')_{135}$ where $(d')_{45}$ and $(d')_{135}$ are the piezoelectric coefficients for the 45° and 135° bars.

† From equations given by Cady, 1946, pp. 88 and 297, we have

$$d' = \pm \left(\frac{\pi^2 C t}{32 l^3 b f_0^2 \rho} \right)^{1/2}$$

where l, b , and t are the length, breadth, and thickness of the bar, ρ its density, and f_0 its frequency in series resonance. C must now be in E.S.U.

the temperature at which the dimensions were measured. At other temperatures, both the density and dimensions are different, but the joint effect of their changes on the capacitance C' is zero for the Y-axis bar, and only 6 parts/million/ $^{\circ}\text{C}$. for the other two bars.

Table 2

Temperature ($^{\circ}\text{C}$.)	25	52.5	80
Y-axis bar :			
C	0.0327 ₈	0.0323 ₇	0.0319 ₈
f_0	74.85 ₁	74.85 ₀	74.84 ₁
$d_{11} = -d'$	6.94 ₈	6.90 ₂	6.85 ₀
Oblique bars : 45° :			
C	0.00453 ₁	0.00455 ₈	0.00457 ₀
f_0	138.59	138.31	138.00
d'	-4.58 ₀	-4.60 ₂	-4.62 ₁
135° :			
C	0.00192 ₄	0.00181 ₀	0.00170 ₀
f_0	174.16	173.76	173.34
d'	-2.37 ₄	-2.30 ₀	-2.24 ₂
Hence	$\begin{cases} d_{11} & 6.95_4 \\ d_{14} & -2.20_6 \end{cases}$	$\begin{cases} 6.91_1 \\ -2.29_3 \end{cases}$	$\begin{cases} 6.86_8 \\ -2.37_9 \end{cases}$
Differences for 27.5°C . :			
d_{11}		$\begin{cases} -0.04_3 \\ -0.04_3 \end{cases}$	$\begin{cases} -0.04_3 \\ -0.04_8 \end{cases}$
d_{14}		-0.08 ₇	-0.08 ₄

Hence the mean temperature coefficients of d_{11} and d_{14} are respectively -2_{80} and $+16_{80}$ parts/million/ $^{\circ}\text{C}$. in the range 25 to 80°C .

These results suggest that any maximum in d_{11} is at a temperature well outside the range 25 to 80°C . (probably at least 100 deg. away).

Although these experiments were not intended as a measurement of the piezoelectric coefficients, the oblique bars give a useful indication of d_{14} ; for, provided only that the orientations of the bars are as stated, any error in d_{14} would occur equally in d_{11} ; so that, since d_{11} determined from these bars agrees with the figure for the Y-axis bar and also with the accepted figure, the value for d_{14} should be equally reliable. The value $(-2.21 \pm 0.1) \times 10^{-8}$ cm/E.S.U. of potential at 25°C ., is higher than most other reported values, and higher than the value -2.0 recommended by Cady (1946, p. 219), though it is lower than the value -2.56 found by Mason (1943).

§ 3. ACKNOWLEDGMENT

Acknowledgment is made to the Engineer-in-Chief of the General Post Office for permission to make use of the information contained in this paper.

REFERENCES

- ANDREEFF, A., FRÉDERICKSZ, V., and KAZARNOWSKY, I., 1929, *Z. Phys.*, **54**, 477.
 CADY, W. G., 1946, *Piezoelectricity* (New York: McGraw-Hill).
 CLAY, J., and KAPER, J. G., 1937, *Physica*, **4**, 311.
 LANGEVIN, A., 1936, *J. Phys. Radium*, sér. 7, **7**, 95.
 LANGEVIN, A., and MOULIN, A. M., 1937, *J. Phys. Radium*, sér. 7, **8**, 257.
 LYNCH, A. C., 1950, *Proc. Phys. Soc. B*, **63**, 323.
 MASON, W. P., 1943, *Bell Syst. Tech. J.*, **22**, 178.
 VAN DYKE, K. S., 1932, *Phys. Rev.*, **42**, 587.

Solvent Effects in Dipole Moment Measurements

By I. G. ROSS* AND R. A. SACK†

* Department of Chemistry, University of Sydney; now at Department of Chemistry, University College, London.

† Division of Electrotechnology, Commonwealth Scientific and Industrial Research Organization, University Grounds, Chippendale, N.S.W.

Communicated by W. E. Duncanson; MS. received 9th May 1950

ABSTRACT. Previous theories advanced to account for the differences between the moments of polar molecules, as measured in dilute solution (in non-polar solvents) and from gas measurements, are discussed on the basis of an internal field function ξ , determining the field inside an ellipsoidal obstacle in a polarized uniform medium. Contours of ξ as a function of the axial ratios of the ellipsoid have been calculated. Onsager's theory of the static dielectric constant is extended to account for the solvent effect in ellipsoidal solute molecules of uniform polarizability, in which the dipole is parallel to one of the axes of the ellipsoid; the new equation thus obtained gives results in better agreement with experimental data than previous theoretical treatments.

The rule that the sign of the solvent effect of a substance should be opposite to that of its Kerr constant is shown not to be of general applicability.

§ 1. INTRODUCTION

A THEORY of the dielectric constant of vapours, liquids and solutions, the molecules of which have permanent electric moments, has been developed by Debye (1929). Although Debye's theory in the case of liquids leads to conclusions which are in qualitative disagreement with experimental facts and has therefore been largely superseded by the later theory of Onsager (1936), it is still usually employed in the determination of dipole moments from measurements on gases or on dilute solutions in non-polar media. For most substances, however, the application of this theory leads to a slight discrepancy between the values of the dipole moments μ_{gas} and μ_{soln} obtained by the two experimental methods. The difference $\Delta\mu = \mu_{\text{soln}} - \mu_{\text{gas}}$, which may be called the solvent effect, is generally negative, but for a minority of substances positive values are found; in either case the magnitude of the solvent effect is dependent on the dielectric constant of the solvent.

It is generally recognized that any theory of the solvent effect must take into account, not only the dielectric constant of the medium, but also particularly the structure of the solute molecules. In theoretical treatments where this factor is taken into consideration a molecule is frequently treated as an ellipsoidal solid with dimensions most closely approximating its actual shape. Such a model is mathematically tractable, and has the further advantage that when placed in a uniform electric field, the field within it is also uniform. Consequently the higher moments in the multipole development of the charge distribution within the molecule are immaterial. The relation between the field inside the ellipsoid and the polarization of the surrounding medium is determined by certain coefficients, hereafter referred to as internal field functions, which depend on the shape of the molecule and its orientation with regard to the field; they are defined in § 2 and evaluated in the Appendix. These internal field functions arise irrespective of whether the problem of the solvent effect is approached from the point of view of Debye's or of Onsager's theory. The former approach has been used by

previous authors; their procedure is critically discussed in § 3. It appears more logical, however, to extend Onsager's theory to the case of dilute solutions; this is done in § 4, and the results obtained are discussed and compared with experiment in § 5. Finally in § 6 a comparison is made between the formulae for the values of the solvent effect and the Kerr constant which shows that the two are not directly related, in contrast to previous statements.

§ 2. THE INTERNAL FIELD FUNCTIONS ξ .

As shown by Maxwell (1873), the disturbance of electric potential produced in a uniform electric field by a small ellipsoidal obstacle of uniform polarizability and with semi-axes a, b, c , depends on the quantities

$$\left. \begin{aligned} \xi_a &= \frac{1}{2} abc \int_0^\infty \frac{d\lambda}{(a^2 + \lambda)^{3/2} (b^2 + \lambda)^{1/2} (c^2 + \lambda)^{1/2}}, \\ \xi_b &= \frac{1}{2} abc \int_0^\infty \frac{d\lambda}{(a^2 + \lambda)^{1/2} (b^2 + \lambda)^{3/2} (c^2 + \lambda)^{1/2}}, \\ \xi_c &= \frac{1}{2} abc \int_0^\infty \frac{d\lambda}{(a^2 + \lambda)^{1/2} (b^2 + \lambda)^{1/2} (c^2 + \lambda)^{3/2}}. \end{aligned} \right\} \dots\dots (1)$$

In the case where we are dealing with an ellipsoidal *cavity* in a uniformly polarized medium, ξ has the simple meaning that the internal field within the cavity is

$$F = E + 4\pi\xi P, \quad \dots\dots (2)$$

where $\xi = \xi_a, \xi_b$, or ξ_c according as the field E is applied parallel to the axis a, b , or c , and P is the polarization. For this reason we may call ξ the *internal field function*.

For arbitrary values of the axes the following identity holds:

$$\xi_a + \xi_b + \xi_c = 1; \quad \dots\dots (3)$$

hence for spherical obstacles all these three quantities (1) become equal to $\frac{1}{3}$ (cf. the Lorentz-Lorenz internal field for a spherical cavity, $F = E + 4\pi P/3$).

Explicit expressions for ξ have been given to date only for the cases where two axes are equal, i.e. for spheroidal obstacles (Maxwell 1873—see Appendix, equations (60), (61)). For unequal values of a, b, c , the formulae (1), despite their simple form, are not suitable for computing ξ as tables of the integrals are not available. General explicit expressions for ξ are derived in the Appendix (equations (46), (54), (58)).

Values of ξ , as a function of the axial ratios, have been evaluated with the use of these formulae, and are shown in Figure 1. In this figure A is the length of the axis parallel to which the field is applied, and B and C are the greater and smaller, respectively, of the other two axes; the dotted lines represent the various limiting spheroidal cases, points on which are evaluated by means of (60) and (61). Values for $A/B > 1$ are given for a short range only, since values in this region may be calculated readily from the identity (3).

§ 3. REVIEW OF PREVIOUS THEORIES OF THE SOLVENT EFFECT

Debye's theory, which assumes that the polar molecules are oriented under the influence of the Lorentz field $F = E + 4\pi P/3$, leads to the well-known formula for the dielectric constant ϵ_{soln} of a liquid mixture

$$\frac{\epsilon_{\text{soln}} - 1}{\epsilon_{\text{soln}} + 2} = \frac{4\pi}{3} \sum N \left(\alpha + \frac{\mu^2}{3kT} \right), \quad \dots\dots (4)$$

where α and μ denote the polarizability and the permanent dipole moment, respectively, of the individual molecules, N their number per unit volume, and the summation extends over all types of molecules present (Debye 1929). To account for the solvent effect, i.e. the difference between the values of μ calculated according to (4) from measurements on dilute solutions and on vapours, a number of empirical relations have been derived, the most satisfactory of which is that due to Goss (1937, 1940); the latter has some theoretical justification, and also takes into account the anisotropy of the medium. In addition the problem has been approached theoretically from two viewpoints.

The first approach is based on the theory of Raman and Krishnan (1928) concerning the refractive index and dielectric constant of polar liquids. This theory considers in detail the internal field acting on each dipole and its deviation from the Lorentz field. It has been adapted to mixtures, especially of polar substances

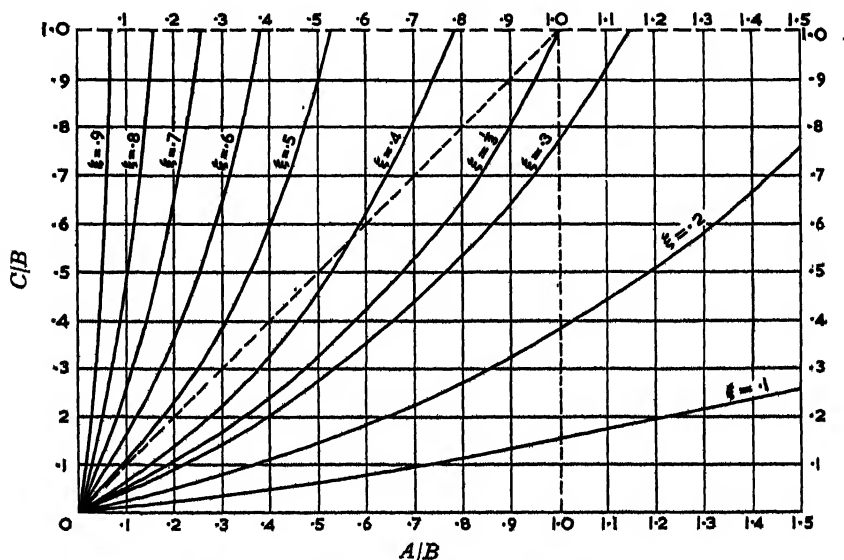


Figure 1. Contours of the internal field function.

with non-polar liquids, by Goss (1934) and other workers. In particular, if small correction terms arising from the anisotropies of the solvent and solute are neglected, then at infinite dilution there is obtained the relation (Le Fèvre and Le Fèvre 1935)

$${}_0P_{\text{soln}}/{}_0P_{\text{gas}} = 1 + 3\Theta(\epsilon - 1)/4\pi\mu^2(\epsilon + 2), \quad \dots\dots(5)$$

where ${}_0P$ is the orientation polarization of the solute, μ its dipole moment, ϵ the dielectric constant of the solvent, and

$$\Theta = \sum_i \mu_i^2 \sigma_i + \sum_{i \neq j} \mu_i \mu_j q_{ij} \quad i, j = 1, 2, 3.$$

Here the μ_i are the components of μ along the principal axes of the ellipsoid representing the molecule, q_{ij} are the components of the polarization field tensor, and $\sigma_i = q_{ii} - 4\pi/3$, i.e. $\sigma_i P$ is the excess of the internal field over the Lorentz-Lorenz internal field for a spherical cavity. For the case where the dipole is parallel to one of the principal axes of the ellipsoid—and for simplicity we shall confine

the discussion to such molecules—and the solvent medium is isotropic, $\Theta = \mu^2 \sigma_i = 4\pi\mu^2 (\xi_i - \frac{1}{3})$, and the expression (5) reduces to

$${}_0P_{\text{soln}}/{}_0P_{\text{gas}} = 1 + 3(\xi - \frac{1}{3})(\epsilon - 1)/(\epsilon + 2)$$

or, since ${}_0P$ is proportional to μ^2

$$S_{\text{RK}} = \frac{\mu_{\text{soln}}}{\mu_{\text{gas}}} = 1 + \frac{3}{2}(\xi + \frac{1}{3}) \frac{\epsilon - 1}{\epsilon + 2} + \dots \quad \dots\dots(6)$$

The Raman-Krishnan theory, both in the simplified form (5) and in the exact form, in which allowance is made for anisotropy effects, has been critically tested by Jenkins and Bauer (1936), who find that the observed values of Θ , as calculated from experimental values of ${}_0P_{\text{soln}}$ in six non-polar solvents and ${}_0P_{\text{gas}}$ may vary within the range of $\pm 50\%$ of the mean value.

An alternative theory, based on a consideration of the moment induced in the solvent by the dissolved polar molecule, has been developed by Frank (1935) and Higasi (1936), of whom the latter especially has been widely quoted in the literature. These authors replace the molecules by a cavity of appropriate shape, with a dipole $\mu = \mu_{\text{gas}}$ at the centre. The field due to this dipole polarizes the surrounding medium, and the volume integral of this polarization, $\Delta\mu$, is taken as the solvent effect. The results obtained by the two authors are inconsistent for two reasons. In the first place, the value obtained for $\Delta\mu$ depends on the shape of the volume of the solvent; thus Higasi, who integrates over a spherical volume, finds, for a spherical molecule, $\Delta\mu/\mu = 0$, while Frank, whose volume is bounded by a surface over which the electric field due to μ has constant magnitude and which resembles a prolate spheroid elongated along the dipole axis, finds $\Delta\mu/\mu = 5\%$ (in benzene); and the values of the solvent effect calculated by the two theories will always differ by this amount. Indeed, even greater differences may be predicted by the consistent use of this approach. The ratio $\Delta\mu/\mu$ becomes larger as the enveloping surface becomes more elongated along the dipole axis, until, in the limiting case of an infinitely long cylindrical surface $\Delta\mu/\mu = (\epsilon - 1)/3\epsilon = 18.6\%$ (in benzene) for spherical solutes. For oblate surfaces, $\Delta\mu/\mu$ is negative, and in the limiting case of a cylindrical medium of infinite radius, with the dipole along the axis of the cylinder, $\Delta\mu/\mu = -37.2\%$.

The reason for this is that even at large distances the contribution to $\int \mathbf{P} d\mathbf{v}$ due to that part of the solvent beyond the chosen boundary surface does not tend to zero. If the volume is assumed spherical and the distribution of polarizable matter isotropic as in the original Clausius-Mosotti-Debye theory, induced polarization does not come into play. The correct approach, on this treatment, is therefore to find the apparent dipole moment *on the Clausius-Mosotti model*, which is that of a spherical unit, composed in this case of the solute molecule, plus the additional solvent contained between it and an enclosing spherical surface (whose radius is immaterial). Thus it appears that Higasi has chosen the correct boundary surface, although, in assuming that $\int \mathbf{P} d\mathbf{v}$, over the volume between his spherical outer boundary 'at infinity' and the actual boundary of the liquid, is zero, his reason for so choosing is incorrect.

Higasi, however, is in error in evaluating $\Delta\mu$. If \mathbf{E} is the field due to μ at a given point, then, following Frank,

$$\mathbf{P} = (\epsilon - 1) \mathbf{E} / 4\pi \quad \text{and} \quad \Delta\mu = (\epsilon - 1) \int \mathbf{E} d\mathbf{v} / 4\pi.$$

According to Higasi, however, $\mathbf{P} = N\alpha\mathbf{E} = (\epsilon - 1) \mathbf{E} / (\epsilon + 2)$ where N is the number of molecules per unit volume and α their polarizability. But the polarizing field at a

point should be the spherical cavity field $F = E + 4\pi P/3$; moreover, in evaluating E , Higasi neglects the influence of the dielectric constant of the intervening medium, and his value for $\Delta\mu$ should be further divided by ϵ on this account. The combined effect of these two factors is to multiply the Higasi expression by $(\epsilon + 2)/3\epsilon$, which leads to the same result as the Frank calculation applied to a spherical volume. Inspection of Higasi's integrals for ellipsoidal cavities shows that its equation can be simply written

$$\mu_{\text{soln}}/\mu_{\text{gas}} = 1 + 3(\xi - \frac{1}{3})(\epsilon - 1)/(\epsilon + 2), \quad \dots\dots(7)$$

or in the amended form,

$$S_H = \mu_{\text{soln}}/\mu_{\text{gas}} = 1 + (\xi - \frac{1}{3})(\epsilon - 1)/\epsilon. \quad \dots\dots(8)$$

As the theory developed below, in which the Clausius-Mosotti approach is not used, leads under certain conditions to formula (7), it is appropriate to draw attention here to the similarity between the Raman-Krishnan (6) and Higasi (7) equations, the latter of which predicts values of $\Delta\mu$ exactly twice those given by the former.

§4. A NEW THEORY OF THE SOLVENT EFFECT

Onsager (1936), in his paper on the static dielectric constant of pure liquids and solutions, in which he considers the polar molecules as being oriented not by the Lorentz field F , but by the cavity field G , carries out the explicit calculations for isotropic spherical molecules only, although he refers to the possible extension to ellipsoidal shapes. Allowance for anisotropy, i.e. for different polarizabilities in different directions within the still spherical molecule, has subsequently been made by Wilson (1939).

We may, on the other hand, take into account the departure from spherical shape of the solute molecules. We treat these as uniform ellipsoids, in which the charge distribution in the case of a vanishing internal field leads to a dipole moment μ in the direction of one of the axes. The ellipsoids are assumed to have an isotropic 'effective' refractive index n , defined, following Wilson, by

$$(n^2 - 1)M/(n^2 + 2)\rho = 4\pi N_A \alpha/3 = {}_D P, \quad \dots\dots(9)$$

where ${}_D P$ is the distortion polarization of the molecule (the sum of the electronic and atomic polarizations), α its mean polarizability, ρ the density of the pure substance at the temperature considered, M its molecular weight and N_A Avogadro's number. Ideally, ${}_D P$ should be the value obtained from gas measurements; if such measurements are not available, one of the usual approximations such as ${}_D P = 1.05$ times $[M_R]_D$, the molecular refraction of the D-line, may be used.

For the treatment of dilute solutions in non-polar media, each ellipsoid is imagined surrounded by an isotropic continuum having the dielectric constant ϵ of the solvent (the limitations of this simplification have been discussed by Onsager (1936)). The external moment μ^* determining the force exerted at large distances on the surrounding liquid is then independent of higher moments in the multipole development of the charge distribution. If G is the cavity field parallel to the dipole under an applied field E in the same direction, Onsager's relation

$$\mu G = \mu^* E \quad \dots\dots(10)$$

must hold irrespective of shape; for it does not matter whether we calculate the interaction energy as the energy of the dipole under the influence of the uniform field, or as the energy of the charges (imagined at large distances) producing that field in the potential of the dipole. This rule still applies if we do not remove the

whole molecule inside its boundary, but, following the procedure of Fröhlich and Sack (1944), merely remove the dipole, leaving the polarizability unaltered. (Equation (10) requires modification if the media are no longer isotropic, for in that case the total interaction energy between the field and the induced polarization within two similar concentric surfaces no longer vanishes.) For a spherical cavity, G is given by

$$G = 3\epsilon E / (2\epsilon + n^2),$$

and in the general case, according to (39), ($\epsilon_0 = n^2$, $\epsilon_1 = \epsilon$)

$$G = \frac{\epsilon}{\epsilon + (n^2 - \epsilon)\xi} E \quad \dots\dots(11)$$

which, with (10), leads to

$$\mu^* = \mu\epsilon / [\epsilon + (n^2 - \epsilon)\xi]. \quad \dots\dots(12)$$

Gas measurements correspond to $\epsilon = 1$, in which case μ^* is the measured gas moment, μ_{gas} ;

$$\mu_{\text{gas}} = \mu / [1 + (n^2 - 1)\xi], \quad \dots\dots(13)$$

whence

$$\frac{\mu^*}{\mu_{\text{gas}}} = \frac{1 + (n^2 - 1)\xi}{\epsilon + (n^2 - \epsilon)\xi} \epsilon. \quad \dots\dots(14)$$

A treatment exactly analogous to Onsager's leads to the same formula as the equation (42) of his paper (1936):

$$\epsilon_{\text{soln}} = n_{\text{soln}}^2 + 4\pi N_s (\mu^*)^2 / 3 kT, \quad \dots\dots(15)$$

where N_s denotes the number of solvent molecules per unit volume. The effective refractive index of the solution, n_{soln} , is obtained by adding the components of all the induced dipoles in the direction of the field; if λ_k denotes the angle between the field and the k th axis of the ellipsoid, and μ_{E}' is the component of the induced moment in the field, then (62), together with (39), (50) and (59), leads to

$$\begin{aligned} (n_{\text{soln}}^2 - \epsilon) E / 4\pi &= N_s [\mu_{\text{E}}']_{\text{av}} \\ &= \frac{1}{3} N_s (n^2 - \epsilon) abc \sum_{k=1}^3 \left[\frac{\cos \lambda_k \cdot E}{\epsilon + (n^2 - \epsilon)\xi_k} \cos \lambda_k \right]_{\text{av}}. \quad \dots\dots(16) \end{aligned}$$

On averaging, $\cos^2 \lambda_k$ has to be replaced by $\frac{1}{3}$, and hence

$$n_{\text{soln}}^2 - \epsilon = \frac{1}{3} v (n^2 - \epsilon) \sum_{k=1}^3 [\epsilon + (n^2 - \epsilon)\xi_k]^{-1}, \quad \dots\dots(17)$$

where v denotes the volume fraction occupied by the solute. Strictly speaking, equation (9) should be replaced by

$$\frac{n^2 - 1}{9} \frac{M}{\rho} \sum_k \frac{1}{1 + (n^2 - 1)\xi_k} = D^P \quad \dots\dots(18)$$

to make the model self-consistent; but the crude nature of some of the approximations involved, in particular the assumption of a uniform homogeneous n^2 , hardly justifies the application of this additional correction.

If the apparent dipole moment, μ_{soln} is determined by applying the Debye equation (4) to the value of ϵ_{soln} calculated with the use of (14), (15) and (17), the generalization of Onsager's formula

$$S_0 = \mu_{\text{soln}} / \mu_{\text{gas}} = 3\epsilon(n^2 + 2) / (n^2 + 2\epsilon)(\epsilon + 2) \quad \dots\dots(19)$$

is found to be

$$S_{RS} = \mu_{\text{soln}} / \mu_{\text{gas}} = \frac{3\epsilon[1 - (n^2 - 1)\xi]}{(\epsilon + 2)[\epsilon + (n^2 - \epsilon)\xi]} \quad \dots\dots (20)$$

or

$$\Delta\mu / \mu_{\text{gas}} = \frac{(\epsilon - 1)[(\epsilon + 2n^2)\xi - \epsilon]}{(\epsilon + 2)[\epsilon + (n^2 - \epsilon)\xi]} \quad \dots\dots (21)$$

For the special case when $\epsilon = n^2$, equation (20) reduces to the uncorrected Higasi formula (7).

§ 5. DISCUSSION AND COMPARISON WITH EXPERIMENTAL RESULTS

If the theory developed above is to be applied to actual substances, the dimensions of the ellipsoidal model representing the molecule must be known. A number of methods by which these dimensions may be estimated have been listed by Jenkins and Bauer (1936). These authors recommend the use of light scattering measurements on the pure (liquid) substance, but in view of the contention of Stuart and Volkmann (1933a) and Mueller (1936) that the classical theory of light scattering cannot be applied to the liquid state, it is preferable to base the choice of axial ratios on the actual dimensions of the molecule. With

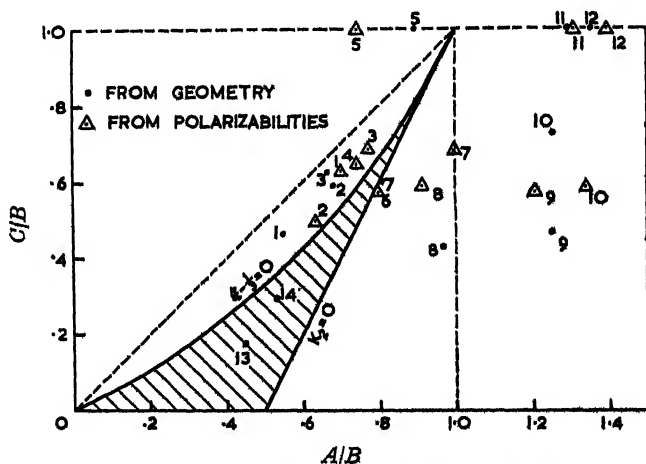


Figure 2. Axial ratios of molecular ellipsoids.

1. Ethyl ether ($10^{18} \times K_1$ at 760 mm. and 5400 Å. = -7.6). 2. sulphur dioxide (-11.1). 3. methyl ether (-5.9). 4. di-isopropyl ketone (-18.4). 5. chloroform (-8.8). 6. diethyl ketone (-7.6). 7. acetone (31.1). 8. pyridine (21.6). 9. chlorobenzene (29.7). 10. nitrobenzene (139). 11. methyl chloride (35.7). 12. methyl bromide (44.3). 13. 3:4:8:9-dibenz-acridine (—). 14. 3:9-dimethylacridine (—).

flexible molecules, however, a variety of molecular shapes is possible, and polarizability measurements must be used, since by the use of such data the most probable molecular configurations can be decided upon. Because the three principal polarizabilities of a molecule are roughly proportional to its linear dimensions in the corresponding directions, the use of these to determine the axial ratios of the ellipsoid gives results not greatly different from those obtained from the geometrical shape of the molecule. To illustrate this, the theory of the solvent effect presented here is tested, using axial ratios calculated by both of these methods, although it should be emphasized that the consideration of the principal polarizabilities is not consistent with the use of an isotropic n^2 in the calculations of § 4. The axial ratios for a number of such molecules are plotted in Figure 2, where the values for

the polarizability ellipsoid have been calculated from data by Stuart and Volkmann (1933 b), and the geometrical ratios have been evaluated using bond radii and angles given by Pauling (1944) and packing radii as given by Rees (1948).

Calculated and Observed Solvent Effects Solutions in Benzene ($\epsilon = 2.226$) at 25° c.

Substance	μ_{gas}	nD	ρ	n^2	Molecular ellipsoid			S_{RS}	S_H	S_{obs}
					A/B	C/B	ξ			
Chlorobenzene	1.69	32.6*	1.101	2.405	G 1.251	0.463	0.181	0.878	0.916	0.90
					P 1.203	0.573	0.215	0.908	0.935	
Nitrobenzene	4.23	33.6*	1.489	3.055	G 1.255	0.727	0.236	0.969	0.947	0.93
					P 1.340	0.585	0.195	0.927	0.924	
Acetone	2.84	16.5*	0.784	1.860	G 0.805	0.600	0.315	0.965	0.990	0.95
					P 0.997	0.679	0.281	0.937	0.971	
Methyl ether	1.29*	14.1*	0.654	1.751	G 0.666	0.628	0.368	0.983	1.019	—
					P 0.771	0.684	0.348	0.967	1.008	
Ethyl ether	1.18	33.2	0.708	1.855	G 0.550	0.462	0.373	0.998	1.022	0.98
					P 0.699	0.628	0.356	0.984	1.013	
Methyl chloride	1.86	15.4	0.906	2.145	G 1.298	1.000	0.267	0.936	0.964	0.91b
					P 1.309	1.000	0.265	0.934	0.963	
Methyl bromide	1.78	19.1	1.674	2.522	G 1.356	1.000	0.257	0.955	0.958	—
					P 1.398	1.000	0.249	0.948	0.954	
Chloroform	1.02°	22.5	1.480	2.160	G 0.892	1.000	0.364	1.021	1.017	1.12
					P 0.741	1.000	0.415	1.065	1.046	
Pyridine	—	25.3*	0.978	2.358	G 0.967	0.427	0.222	0.912	0.939	—
					P 0.912	0.487	0.253	0.940	0.956	
Sulphur dioxide	1.62d	10.9	1.369	1.911	G 0.681	0.590	0.354	0.988	1.012	1.00d
					P 0.634	0.497	0.349	0.984	1.009	

G and P denote that the dimensions of the ellipsoidal model have been calculated from geometry and from polarizabilities, respectively. The observed values of μ_{gas} , nD and S have been taken from (a) Groves and Sugden (1937), (b) Barclay and Le Fèvre (1950), (c) Goss (1940), (d) Le Fèvre and Ross (1950); the remainder are from the Appendix to *Trans. Faraday Soc.*, 1934, 30, and Böttcher (1936). The values of nD are the observed gas values where these are reliable; the others, indicated by an asterisk, are calculated as $1.05[M_R]D$. Densities are from *International Critical Tables*, Vol. 3, except for methyl bromide, estimated from the value (1.7116) at 0°.

The results obtained on applying the new theory to various substances are summarized in the Table; the calculated values, S_{RS} , of $\mu_{\text{soln}}/\mu_{\text{gas}}$ are compared with the observed values, S_{obs} , where reasonably reliable experimental results are available (it should be mentioned here that most of the experimental figures S_{obs} have an accuracy not much better than within 1–2%). Also tabulated, for comparison, are the results S_H predicted by the use of the amended Higasi equation (8).

It is seen that rather better agreement between theory and experiment is obtained from the new equation (20) than from (8), and that on the whole values of S calculated from polarizability data give better results than the geometrical values. The theory successfully predicts a small negative $\Delta\mu$ for sulphur dioxide and diethyl ether; methyl chloride, and especially chloroform, show solvent effects much greater than those calculable by any of the methods used.

On differentiating (20) with respect to ϵ , it is seen that the sign of $\partial S/\partial\epsilon$ is determined by the expression $(\epsilon^2 + 2n^2)\xi - \epsilon^2$. If $(1 + 2n^2)\xi < 1$, then the apparent moment should always decrease with increase in ϵ ; if $(1 + 2n^2)\xi > 1$, the equation predicts an increase in μ_{soln} up to a maximum when $(\epsilon^2 + 2n^2)\xi = \epsilon^2$, followed by a decrease; for almost all substances this maximum will be reached at values of ϵ less than those of the normally available non-polar solvents, and hence μ_{soln} should decrease with increasing ϵ of solvent.

This has long been recognized with substances with negative $\Delta\mu$, but is contrary to the general belief when $\Delta\mu$ is positive; however, available data on the variation of the solvent effect with dielectric constant of the solvent are not accurate or numerous enough to support the general assumption; indeed, the results obtained for chloroform and bromoform by Müller (1933), who first drew attention to the solvent effect, if anything support our conclusion.

In view of the failure of Debye's theory to account correctly for the properties of polar liquids it may be pertinent to ask if the use of equation (4) should not be abandoned altogether. The equations (14) and (15) give a much more straightforward correlation between the molecular dipole moments and the macroscopic dielectric constant. Although the theory in its present state only applies to dilute solutions in non-polar media, with the additional assumption that the volumes are additive and that no strong short-range interaction is present between the molecules, an extension based on Onsager's approach appears to be promising enough to render unnecessary a re-interpretation of apparent dipole moments figuring in Debye's formula (4), such as given by equation (20).

§ 6. THE KERR CONSTANT RULE

A final point concerns the rule first stated by Goss (1934), and since frequently reiterated (cf. a review by Smith 1948), that the sign of the solvent effect, which according to the older theories is determined by $\xi - \frac{1}{3}$, should be opposite to that of the Kerr constant of the substance concerned. In fact this is not generally so, and the magnitude of the Kerr constant need bear no simple relation to the magnitude of the solvent effect.

The Kerr constant K of a molecule, a measure of its electric birefringence, is the sum of two terms K_1 and K_2 ; K_1 , which depends on the molecular anisotropy, is positive and generally small; K_2 , which involves the dipole moment as well, is generally very much larger than K_1 , except for molecules with small or zero moments. The three principal polarizabilities of the molecule, α , β , γ , define its polarizability ellipsoid; if μ is parallel to one of the axes of this ellipsoid, say α , then K_2 is proportional to $\mu^2(2\alpha - \beta - \gamma)$; if α is the greatest or least axis of the ellipsoid, then K_2 will be respectively positive or negative; if α is the mid-axis, then K_2 may have either sign, and will be zero if $\alpha = \frac{1}{2}(\beta + \gamma)$.

If we assume that α , β , γ are proportional to the axes A , B , C of the geometrical ellipsoid (and it is on this not very accurate assumption that the rule is based), then $K_2 \sim \mu^2(2A - B - C)$, and we may attempt to correlate K_2 and $\xi - \frac{1}{3}$. This is done in Figure 2, in which the relation $K_2 = 0$ is represented by the heavy straight line, and $\xi - \frac{1}{3} = 0$ by the heavy curved line. Inspection of the figure shows that, for molecules lying in the region between these lines, shown shaded in Figure 2, K_2 and $\xi - \frac{1}{3}$ have the same (negative) sign. The contribution of the anisotropy term K_1 is usually small, and cannot be evaluated with reference to the polarizabilities alone; being positive, its effect will be that certain molecules lying close to, but to the left of, the line $K_2 = 0$, may yet have positive overall Kerr constants $K = K_1 + K_2$, but the general conclusion that there can arise cases in which K and $(\xi - \frac{1}{3})$ have the same sign, still stands, contrary to Goss' rule. In Figure 2 there are included also two molecules whose geometrical dimensions suggest that they might fail to comply with this Kerr constant rule. Further, it will be observed that, especially for molecules in which μ is parallel to the mid-

axis, a large value of K does not necessarily imply a large value of $\xi - \frac{1}{2}$, and hence of the solvent effect.

With the use of the new theory the sign of the solvent effect according to (21) is determined by the expression $(\epsilon + 2n^2)\xi - \epsilon$; ξ alone is thus not sufficient to determine the sign of $\Delta\mu$, and the connection between $\Delta\mu$ and the Kerr constant becomes even more remote. The Table and Figure 2 show that sulphur dioxide and methyl chloride both show effects which are contrary to Goss' rule.

ACKNOWLEDGMENT

The authors wish to thank Professor R. J. W. Le Fèvre for helpful discussions on this topic and for permission to use unpublished results.

APPENDIX

CALCULATION OF THE FUNCTIONS ξ

In §2 it is pointed out that the general equations (1) for the internal field functions ξ are inconvenient for deriving numerical values for ξ in the case of ellipsoids with three unequal axes. A more convenient approach is found in the theory of ellipsoidal harmonics (Lamé functions) in association with Jacobian elliptic functions (cf., for example, Whittaker and Watson 1927).

If $a > b > c$, the appropriate modulus k , complementary modulus k' , and base length l are

$$k = \left(\frac{a^2 - b^2}{a^2 - c^2} \right)^{1/2}; \quad k' = (1 - k^2)^{1/2} = \left(\frac{b^2 - c^2}{a^2 - c^2} \right)^{1/2}; \quad l = (a^2 - c^2)^{1/2} \quad \dots\dots (22)$$

The Jacobian elliptic functions

$$\text{sn } u = \text{sn}(u, k), \quad \text{cn } u = \text{cn}(u, k), \quad \text{dn } u = \text{dn}(u, k) \quad \dots\dots (23)$$

are introduced, defined by

$$\left(\frac{d \text{sn } u}{du} \right)^2 = (1 - \text{sn}^2 u) (1 - k^2 \text{sn}^2 u); \quad \text{sn } 0 = 0; \quad \left(\frac{d \text{sn } u}{du} \right)_{u=0} = 1 \quad \dots\dots (24)$$

$$\text{and} \quad \text{cn}^2 u = 1 - \text{sn}^2 u, \quad \text{dn}^2 u = 1 - k^2 \text{sn}^2 u; \quad \text{cn } 0 = \text{dn } 0 = 1. \quad \dots\dots (25)$$

The complete elliptic integrals of the first kind

$$\left. \begin{aligned} K &= K(k) = \int_0^{\pi/2} (1 - k^2 \sin^2 \phi)^{-1/2} d\phi, \\ iK' &= i \int_0^{\pi/2} (1 - k'^2 \sin^2 \phi)^{-1/2} d\phi, \end{aligned} \right\} \quad \dots\dots (26)$$

are the semi- or quarter-periods of the doubly periodic functions (23); the values of the functions for special values of u are

$$\text{sn } K = 1, \quad \text{cn } K = 0, \quad \text{dn } K = k' \quad \dots\dots (27)$$

$$\text{sn}(K + iK') = k^{-1}; \quad \text{cn}(K + iK') = -ik'/k; \quad \text{dn}(K + iK') = 0; \quad \dots\dots (28)$$

whereas at $u = iK'$ all three functions have simple poles with residues k^{-1} , $-ik^{-1}$, $-i$ respectively.

The Cartesian coordinates x, y, z can now be expressed in the form

$$\left. \begin{aligned} x &= k^2 l \text{sn } \alpha \text{sn } \beta \text{sn } \gamma, \\ y &= -(k^2/k') l \text{cn } \alpha \text{cn } \beta \text{cn } \gamma, \\ z &= (i/k') l \text{dn } \alpha \text{dn } \beta \text{dn } \gamma. \end{aligned} \right\} \quad \dots\dots (29)$$

The surfaces upon which any one of the variables α , β , γ are constant are confocal quadrics; they are ellipsoids with semi-axes

$$a = kl \operatorname{sn} \alpha, \quad b = ikl \operatorname{cn} \alpha, \quad c = il \operatorname{dn} \alpha \quad \dots\dots(30)$$

for constant α lying between iK' and $K + iK'$, hyperboloids of one sheet for β lying between K and $K + 2iK'$, and hyperboloids of two sheets for γ lying between 0 and $4K$.

A harmonic function V , satisfying

$$\nabla^2 V = 0 \quad \dots\dots(31)$$

can be separated into Lamé functions $V = \Lambda(\alpha)M(\beta)N(\gamma)$ where each of the functions satisfies an equation of the type

$$d^2\Lambda/d\alpha^2 = [n(n+1)k^2 \operatorname{sn}^2 \alpha + A]\Lambda, \quad \dots\dots(32)$$

where n is the degree of the harmonic, and A a separation constant. In the case of a homogeneous field, $n=1$, and Λ is a multiple of $\operatorname{sn} \alpha$, $\operatorname{cn} \alpha$, or $\operatorname{dn} \alpha$, depending on the direction of the field. The introduction into a medium of dielectric constant ϵ_1 of an ellipsoid of dielectric constant ϵ_0 whose surface is given by the relation $\alpha = \alpha_0$ will introduce a disturbance depending mathematically upon the second solution of (32), viz.

$$\Lambda(\alpha) \int_{iK'}^{\alpha} \frac{du}{[\Lambda(u)]^2},$$

the lower limit being chosen as the pole of Λ as this corresponds to a falling off of the disturbance at large distances ($\alpha \rightarrow iK'$). Taking the case of the field acting in the direction of the *longest* axis, the potentials outside and inside the ellipsoid, V_0 and V_1 , can be written

$$\left. \begin{aligned} V_0 &= -k^2 l (E - k'^2 H_a) \int_{iK'}^{\alpha} \frac{du}{\operatorname{sn}^2 u} \operatorname{sn} \alpha \operatorname{sn} \beta \operatorname{sn} \gamma, \\ V_1 &= -k^2 l G_a \operatorname{sn} \alpha \operatorname{sn} \beta \operatorname{sn} \gamma, \end{aligned} \right\} \quad \dots\dots(33)$$

where G_a represents the cavity field in the ellipsoid, and $\frac{1}{2}H_a$ is a measure of the effective induced dipole. The quantities E , G_a , and H_a are related by the continuity conditions

$$V \text{ continuous at } \alpha = \alpha_0, \quad \dots\dots(34)$$

$$(\epsilon \operatorname{grad} V)_{\text{normal}} \text{ continuous at } \alpha = \alpha_0, \quad \dots\dots(35)$$

the latter relation being equivalent to

$$\epsilon \partial V / \partial \alpha \text{ continuous at } \alpha = \alpha_0. \quad \dots\dots(36)$$

Substituting (34) and (36) into (33), we obtain

$$E - k'^2 H_a \int_{iK'}^{\alpha_0} \frac{du}{\operatorname{sn}^2 u} - G_a = 0 \quad \dots\dots(37)$$

$$\left[\epsilon_1 (E - k'^2 H_a \int_{iK'}^{\alpha_0} \frac{du}{\operatorname{sn}^2 u}) - \epsilon_0 G_a \right] d \operatorname{sn} \alpha_0 / d \alpha_0 - \epsilon_1 H_a / k^2 \operatorname{sn} \alpha_0 = 0, \quad \dots\dots(38)$$

and hence

$$G_a = \epsilon_1 E / [\epsilon_1 - (\epsilon_1 - \epsilon_0) \xi_a], \quad \dots\dots(39)$$

$$H_a = G_a k^2 (\epsilon_1 - \epsilon_0) \operatorname{sn} \alpha_0 (d \operatorname{sn} \alpha_0 / d \alpha_0) / \epsilon_1, \quad \dots\dots(40)$$

where
$$\xi_a = -\operatorname{sn} \alpha_0 \left(d \operatorname{sn} \alpha_0 / d\alpha_0 \right) \int_{iK'}^{\alpha_0} du / \operatorname{sn}^2 u. \quad \dots\dots(41)$$

The difficulties arising from the complex values of α_0 can be eliminated by the transformation

$$\alpha' = \alpha_0 - iK' \quad \dots\dots(42)$$

with (cf. Whittaker and Watson 1927)

$$\operatorname{sn} \alpha_0 = 1 / (k \operatorname{sn} \alpha'). \quad \dots\dots(43)$$

The equations (40) and (41) can then be expressed as

$$\begin{aligned} \xi_a &= -\frac{1}{\operatorname{sn} \alpha'} \frac{d}{d\alpha'} \left(\frac{1}{\operatorname{sn} \alpha'} \right) \int_0^{\alpha'} \operatorname{sn}^2 u \, du \\ &= \frac{1}{k^2} \frac{\operatorname{cn} \alpha' \operatorname{dn} \alpha'}{\operatorname{sn}^3 \alpha'} \int_0^{\alpha'} (1 - \operatorname{dn}^2 u) \, du \\ &= \frac{1}{k^2} \frac{\operatorname{cn} \alpha' \operatorname{dn} \alpha'}{\operatorname{sn}^3 \alpha'} \left(\alpha' - \int_0^{\alpha'} \operatorname{dn}^2 u \, du \right), \quad \dots\dots(44) \end{aligned}$$

and
$$H_a = \frac{\epsilon_0 - \epsilon_1}{\epsilon_1} \frac{\operatorname{cn} \alpha' \operatorname{dn} \alpha'}{\operatorname{sn}^3 \alpha'} G_a. \quad \dots\dots(45)$$

Introducing the angle $\theta = \sin^{-1}(\operatorname{sn} \alpha')$, (44) and (45) become

$$\xi_a = \frac{1}{k^2} \frac{\cos \theta (1 - k^2 \sin^2 \theta)^{1/2}}{\sin^3 \theta} [F(k, \theta) - E(k, \theta)], \quad \dots\dots(46)$$

$$H_a = G_a \frac{\cos \theta (1 - k^2 \sin^2 \theta)^{1/2}}{\sin^3 \theta} \frac{\epsilon_0 - \epsilon_1}{\epsilon_1}, \quad \dots\dots(47)$$

where $F(k, \theta)$ and $E(k, \theta)$ represent the elliptic integrals of the first and second kinds, respectively, and

$$\cos \theta = \operatorname{cn} \alpha' = i \operatorname{dn} \alpha_0 / k \operatorname{sn} \alpha_0 = c/a. \quad \dots\dots(48)$$

If the applied field is in the direction of the mid-axis, $2b$, of the ellipsoid, then the following equations replace equations (33) for the external and internal potentials:

$$\left. \begin{aligned} V_e &= (k^2/k') l (E + k^{-2} H_b) \int_{iK'}^{\alpha_0} du / \operatorname{cn}^2 u \operatorname{cn} \alpha \operatorname{cn} \beta \operatorname{cn} \gamma, \\ V_i &= (k^2/k') l G_b \operatorname{cn} \alpha \operatorname{cn} \beta \operatorname{cn} \gamma. \end{aligned} \right\} \quad \dots\dots(49)$$

From the continuity conditions (34) and (36) we obtain

$$G_b = \frac{\epsilon_1 E}{\epsilon_1 - (\epsilon_1 - \epsilon_0) \xi_b}; \quad H_b = \frac{\epsilon_0 - \epsilon_1}{\epsilon_1} \operatorname{cn} \alpha_0 \frac{d \operatorname{cn} \alpha_0}{d \alpha_0} k^2 G_b, \quad \dots\dots(50)$$

where
$$\xi_b = -\operatorname{cn} \alpha_0 \frac{d \operatorname{cn} \alpha_0}{d \alpha_0} \int_{iK'}^{\alpha_0} du / \operatorname{cn}^2 u. \quad \dots\dots(51)$$

On substituting
$$\alpha_0 = K + iK' - \alpha'' \quad \dots\dots(52)$$

with
$$\operatorname{sn} \alpha_0 = \frac{\operatorname{dn} \alpha''}{k \operatorname{cn} \alpha''}; \quad \operatorname{cn} \alpha_0 = \frac{k'}{i k \operatorname{cn} \alpha''}; \quad \operatorname{dn} \alpha_0 = \frac{k' \operatorname{sn} \alpha''}{i \operatorname{cn} \alpha''} \quad \dots\dots(53)$$

we obtain for (51)

$$\begin{aligned}\xi_b &= \frac{1}{\text{cn } \alpha''} \frac{d}{d\alpha''} \left(\frac{1}{\text{cn } \alpha''} \right) \int_{\alpha''}^K \text{cn}^2 u \, du = \frac{\text{sn } \alpha'' \, \text{dn } \alpha''}{k^2 \text{cn}^3 \alpha''} \int_{\alpha''}^K (\text{dn}^2 u - k'^2) \, du \\ &= \frac{\sin \phi (1 - k^2 \sin^2 \phi)^{1/2}}{k^2 \cos^3 \phi} [E(k) - E(k, \phi) - k'^2 K(k) + k'^2 F(k, \phi)], \quad \dots (54)\end{aligned}$$

where $\sin \phi = \text{sn } \alpha'' = \text{dn } \alpha_0 / k \text{cn } \alpha_0 = c/b$, $\dots (55)$

and $E(k) = E(k, \pi/2)$ represents the complete elliptic integral of the second kind. For (50) we obtain from (53) and (55)

$$H_b = \frac{\epsilon_0 - \epsilon_1}{\epsilon_1} \frac{\text{sn } \alpha'' \, \text{dn } \alpha''}{\text{cn}^3 \alpha''} k'^2 G_b = \frac{\epsilon_0 - \epsilon_1}{\epsilon_1} \frac{\sin \phi (1 - k^2 \sin^2 \phi)^{1/2}}{\cos^3 \phi} k'^2 G_b. \quad \dots (56)$$

Finally, if the external field is in the direction of the minor axis, $2c$, of the ellipsoid, the equations equivalent to (33) and (49) are

$$\left. \begin{aligned}V_0 &= -(i/k') \left[E + H_0 \int_{iK'}^{\alpha} du / \text{dn}^2 u \right] \text{dn } \alpha \, \text{dn } \beta \, \text{dn } \gamma, \\ V_1 &= -(i/k') G_0 \, \text{dn } \alpha \, \text{dn } \beta \, \text{dn } \gamma,\end{aligned} \right\} \quad \dots (57)$$

which, with the boundary conditions (34) and (36), and the substitutions (52), (53) and (55), leads to

$$\begin{aligned}\xi_c &= \frac{\text{sn } \alpha''}{\text{cn } \alpha''} \frac{d}{d\alpha''} \left(\frac{\text{sn } \alpha''}{\text{cn } \alpha''} \right) \int_{\alpha''}^K \frac{\text{cn}^2 u}{\text{sn}^2 u} \, du \\ &= \frac{\text{sn } \alpha'' \, \text{dn } \alpha''}{\text{cn}^3 \alpha''} \int_{\alpha''}^K \left[-\text{dn}^2 u - \frac{d}{du} \left(\frac{\text{cn } u \, \text{dn } u}{\text{sn } u} \right) \right] du = \frac{\text{dn}^2 \alpha''}{\text{cn}^2 \alpha''} - \frac{\text{sn } \alpha'' \, \text{dn } \alpha''}{\text{cn}^3 \alpha''} \\ &\quad \times \int_{\alpha''}^K \text{dn}^2 u \, du = \frac{1 - k^2 \sin^2 \phi}{\cos^2 \phi} - \frac{\sin \phi (1 - k^2 \sin^2 \phi)^{1/2}}{\cos^3 \phi} [E(k) - E(k, \phi)].\end{aligned} \quad \dots (58)$$

Also

$$G_c = \frac{\epsilon_1 E}{\epsilon_1 - (\epsilon_1 - \epsilon_0) \xi_c}; \quad H_0 = k'^2 \frac{\epsilon_0 - \epsilon_1}{\epsilon_1} \frac{\sin \phi (1 - k^2 \sin^2 \phi)^{1/2}}{\cos^3 \phi} G_0. \quad \dots (59)$$

Alternatively, one of the quantities ξ_a , ξ_b , ξ_c could have been derived from the other two by the use of (3).

Finally, to complete the treatment, we restate here the formulae for ξ for spheroidal molecules (Maxwell 1873).

(i) Prolate spheroid, $a > b = c$, $e = (1 - b^2/a^2)^{1/2}$,

$$\xi_a = \left(\frac{1}{e^2} - 1 \right) \left(\frac{1}{2e} \ln \frac{1+e}{1-e} - 1 \right), \quad \xi_b = \xi_c = \frac{1}{2}(1 - \xi_a). \quad \dots (60)$$

(ii) Oblate spheroid, $a = b > c$, $e = (1 - c^2/a^2)^{1/2}$,

$$\xi_c = 1/e^2 - (1 - e^2)^{1/2} \sin^{-1} e / e^3, \quad \xi_a = \xi_b = \frac{1}{2}(1 - \xi_c). \quad \dots (61)$$

The obstacle produces a field at large distances which, with the use of (29), (30), (33), (47), (49), (56), (57) and (59) is found equal to that of a dipole of strength

$$\mu' = \frac{1}{3} (\epsilon_0 - \epsilon_1) abc G$$

in the direction of the applied field, in agreement with Maxwell's results.

The values of ξ_a , ξ_b , ξ_c as functions of the ratios $a : b : c$ have been calculated by first evaluating them for given values of k and ϕ (or θ), using four-figure tables of the circular functions and elliptic integrals, and interpolating for definite values of ξ (0.1, 0.2, . . . , 0.9). The interpolated values of k and ϕ (or θ) were then used to calculate the corresponding ratios $a : b : c$, and the results are shown in Figure 1. The accuracy obtained does not justify the presentation of the results in tabular form.

REFERENCES

- BARCLAY, G., and LE FÈVRE, R. J. W., 1950, *J. Chem. Soc.*, 556.
 BÖTTCHER, C. J. F., 1936, *Physica*, 6, 59.
 DEBYE, P., 1929, *Polar Molecules* (New York : The Chemical Catalog Co.), Chapter 3.
 FRANK, F. C., 1935, *Proc. Roy. Soc. A*, 152, 171.
 FRÖHLICH, H., and SACK, R., 1944, *Proc. Roy. Soc. A*, 182, 388.
 GOSS, F. R., 1934, *J. Chem. Soc.*, 696; 1937, *Ibid.*, 1915; 1940, *Ibid.*, 752.
 GROVES, L. G., and SUGDEN, S., 1937, *J. Chem. Soc.*, 1779.
 HIGASI, K., 1936, *Sci. Pap. Inst. Phys. Chem. Res.*, Tokyo, 28, 284.
 JENKINS, H. O., and BAUER, S. H., 1936, *J. Amer. Chem. Soc.*, 58, 2435.
 LE FÈVRE, C. G., and LE FÈVRE, R. J. W., 1935, *J. Chem. Soc.*, 1747.
 LE FÈVRE, R. J. W., and ROSS, I. G., 1950, *J. Chem. Soc.*, 283.
 MAXWELL, J. C., 1873, *Electricity and Magnetism* (Oxford : Clarendon Press), Section 437.
 MUELLER, HANS, 1936, *Phys. Rev.*, 50, 547.
 MÜLLER, HORST, 1933, *Phys. Z.*, 34, 689.
 ONSAGER, L., 1936, *J. Amer. Chem. Soc.*, 58, 1486.
 PAULING, L., 1944, *Nature of the Chemical Bond* (Cornell : University Press).
 RAMAN, C. V., and KRISHNAN, K. S., 1928, *Proc. Roy. Soc. A*, 117, 589.
 REES, A. L. G., 1948, *J. Chem. Phys.*, 16, 995.
 SMITH, J. W., 1948, *Science Prog.*, 36, 483.
 STUART, H. A., and VOLKMAN, H., 1933 a, *Z. Phys.*, 83, 461; 1933 b, *Ann. Phys., Lpz.*, [v], 18, 121.
 WHITTAKER, E. T., and WATSON, G. N., 1927, *Modern Analysis* (Cambridge : University Press), Chapters 22, 23.
 WILSON, J. N., 1939, *Chemical Reviews*, 25, 377.

A Study of the Horizontal Irregularities of the Ionosphere

BY B. H. BRIGGS AND G. J. PHILLIPS

Cavendish Laboratory, Cambridge

Communicated by J. A. Ratcliffe; MS. received 5th April 1950

ABSTRACT. The theory of diffraction by a random screen developed by Booker, Ratcliffe and Shinn is presented in a convenient form for practical application in ionospheric experiments. It is shown that measurements of the correlation of the fading of the reflected wave observed at spaced receiving points can be used to find the extent of the angular spreading of the downcoming wave.

Histograms are given to show the frequency of occurrence of different degrees of angular spreading observed during a series of experiments using pulse transmissions at vertical incidence.

For a frequency of 2.4 Mc/s. it is most common to find that the downcoming wave has an angular spread such that the amplitude falls to half value at an angle of 5° for regions E and F. For region F observed on 4.8 Mc/s., the corresponding value is 2.5° . There is no evidence for any pronounced seasonal or diurnal variations.

§ 1. INTRODUCTION

WHEN a radio wave is returned from the ionosphere, the amplitude of the reflected wave is found to vary in a random way, and the wave is said to 'fade'. This fading is often dissimilar at two receiving points spaced about one wavelength apart on the ground (Ratcliffe and Pawsey 1933, Pawsey 1935). From this fact it may be deduced that the reflected wave must consist of a cone of rays spread over a considerable range of angles. In this paper we are concerned only with reflection at normal incidence. For this case, simple arguments used by Ratcliffe and Pawsey showed that the deviation of the extreme rays was of the order of 15° from the vertical for the normal E and F layers, and somewhat greater for the abnormal-E layer. The methods used, however, gave only the order of magnitude of the effects, because there was at the time no statistical theory which could be used to make accurate deductions from the observations. Such a theory has recently been developed by Booker, Ratcliffe and Shinn (1950) (we will refer to this paper as B.R.S.). These authors have considered the theoretical problem of diffraction by a random screen such as the ionosphere, using the auto-correlation functions to describe the properties of the random variables which arise.

The first part of the present paper is theoretical in nature, and aims at presenting the treatment of B.R.S. in a form which is convenient for practical applications. In § 2, we first summarize some of the results obtained by these authors. These results apply to the case of a plane wave incident normally on a random diffracting screen, which varies in one direction only—a 'one-dimensional' screen. In § 3, a standard form is assumed for the function which gives distribution of received energy as a function of the angle from the normal to the screen. In this way distributions of varying 'width' are represented by a single analytical expression with one variable parameter. The expected correlation between the amplitude at two points at a given distance apart is evaluated for this case, using the methods of B.R.S. An alternative method of calculation is then given, and is shown to give the same results. In § 4 the new method is used to extend the analysis to

cover the case in which a spherical wave from a point source falls on a two-dimensional random screen. These results are directly applicable to the ionospheric experiments.

In §5 it is shown that the angular spread of received energy can be deduced from the phase differences at spaced receiving points, instead of from the amplitude differences discussed earlier. Now the apparent direction of arrival of the reflected wave as indicated by a radio direction finder is directly related to these phase differences, and so the average departure of this direction from the mean direction may be related to the angular spread of energy in the reflected wave.

In §6 we show how, subject to certain definitions and assumptions, it is possible to deduce a 'size' for the irregularities in the screen which are responsible for the diffraction process. The advantage of this method of expressing the results is that the size of the irregularities as defined in §6 is theoretically independent of wavelength for a diffracting screen of given characteristics, unlike the angular spectrum which, for a given screen, has a width inversely proportional to the frequency.

In §7 experimental results are given. These are obtained entirely from measurements of the correlation of signal amplitude. The experimental arrangement is one in which the fading of an echo returned from the ionosphere is observed at three receiving points placed at the corners of a right-angled triangle as described by Mitra (1949).

Observations on echoes from Regions E, F, and E_s have been made on many occasions (both day and night) over a period of one year, and statistical data are given for each region.

The variation of the angular spectrum with frequency has been verified by comparing the statistical results obtained on two frequencies, 2.4 Mc/s. and 4.8 Mc/s., and we have also shown that the size of the irregularities is, on the average, the same on the two frequencies.

§2. SUMMARY OF THEORETICAL TREATMENT OF A ONE-DIMENSIONAL RANDOM SCREEN

In this section we summarize some of the results obtained by B.R.S.

Consider a plane wave falling on a random screen which is one-dimensional, i.e. varying only in the X -direction.

The incident wave may be polarized with its electric vector either along or at right angles to the X -axis (Figures 1(a) and (b)). The two cases may be treated in a symmetrical manner by working in terms of the magnetic field component H_y of the reflected wave in the first case, and the electric field component E_y in the second case.

These components may be expressed as the sum of contributions from an angular spectrum* of plane waves by means of the following equations in which $s \equiv \sin \theta$:

$$H_y = \eta \int_{-\infty}^{\infty} A_1(s) \exp \left(2\pi i \frac{sx - cz}{\lambda} \right) \frac{ds}{\lambda}, \quad \dots (1)$$

$$E_y = \int_{-\infty}^{\infty} A_2(s) \exp \left(2\pi i \frac{sx - cz}{\lambda} \right) \frac{ds}{\lambda}, \quad \dots (2)$$

* The spectral function employed here is slightly different from that in B.R.S., so that $A_1(s) \cos \theta$ would be equivalent to their $P(s)$.

where η is the ratio of magnetic to electric field intensities for a plane wave in free space.

To find the distribution of power in the angular spectrum consider the power received in a detecting device which responds to a small range of angles s to $s + \Delta s$.

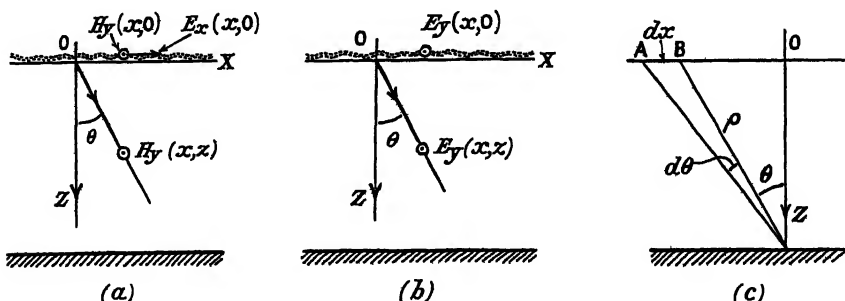


Figure 1. Diagrams to show the notation used for the field components and the geometry of a random screen which varies only in the direction of the X-axis.

We have to perform the above integrations between the limits s and $s + \Delta s$. Assuming this range is large enough for the functions $A_1(s)$ and $A_2(s)$ to perform many random fluctuations the integration amounts to the summation of a number of complex quantities whose phases are random. Now we are not interested in the precise amplitude and phase of the field intensities for a given case. We wish, rather, to know their magnitude as a time average over a number of successive occasions for which the spectrum functions may have changed in detail but not in statistical properties. Taking such an average (which we shall denote by a bar) it can be shown that for the range s to $s + \Delta s$

$$\overline{|H_y|^2} \propto \overline{|A_1(s)|^2} \Delta s,$$

$$\overline{|E_y|^2} \propto \overline{|A_2(s)|^2} \Delta s.$$

The power flux in the direction $s = \sin \theta$ is therefore proportional to $\overline{|A_1(s)|^2} \Delta s$ and $\overline{|A_2(s)|^2} \Delta s$ respectively. Alternatively, if we use a detecting device which accepts a small range of angles θ to $\theta + d\theta$ we will receive powers $\overline{|A_1(s)|^2} \cos \theta d\theta$ and $\overline{|A_2(s)|^2} \cos \theta d\theta$ in the two cases.

Now consider a two-dimensional screen illuminated by a linearly polarized plane wave. It seems reasonable to require that for a random screen, the power should be scattered similarly in the two planes in and at right angles to the direction of polarization. This requires that $\overline{|A_1(s)|^2} = \overline{|A_2(s)|^2}$. We will assume that the screen satisfies this condition, and in future either function will be denoted by $\overline{|A(s)|^2}$.

If in equations (1) and (2) we put $z=0$, we get the variation of the field components of the reflected wave just after it leaves the screen. These fluctuations must be characteristic of the irregularities in the screen itself. Further, $A(s)$ is simply the Fourier transform of the appropriate field component. It follows that the finer the structure in the screen, the wider the 'spread' of the function $A(s)$. Irregularities in the screen of the order of one wavelength correspond to $s=1$, and give rise to energy travelling at 90° to the normal. Irregularities smaller than this give $|s| > 1$ and give rise to evanescent waves

which can be neglected at distances of more than several wavelengths from the screen.

In the ionospheric experiments it is possible to observe the *amplitude* of a field component in a horizontal plane ($z = \text{constant}$) as a function of x . A theorem due to Uhlenbeck relates the auto-correlation function of the fluctuation of amplitude of a random *time* variable to the Fourier transform of the power spectrum. With an appropriate change of variable, the same theorem gives from (1) or (2)

$$\rho_R(\xi) = \left[\frac{\int_{-1}^{+1} |A(s)|^2 \cos(2\pi s\xi) ds}{\int_{-1}^{+1} |A(s)|^2 ds} \right]^2 \quad \dots\dots (3)^*$$

where $\rho_R(\xi)$ is defined by the equation

$$\rho_R(\xi) = \frac{\overline{R(x)R(x+\xi)} - [\overline{R(x)}]^2}{[\overline{R(x)}]^2 - [\overline{R(x)}]^2} \quad \dots\dots (4)$$

R is the amplitude of the field component E_y or H_y and the bars denote averaging over x . The function $\rho_R(\xi)$ measures the average correlation between values of R measured at a distance $\lambda\xi$ apart (i.e. ξ is the distance measured in terms of the wavelength as unit).

By placing the limits in the integrals in (3) equal to ± 1 rather than $\pm \infty$ we have omitted the contribution of any evanescent waves in the angular spectrum. This result holds in general for observations made at a distance of many wavelengths from the screen. It holds at all distances if no evanescent waves are present, i.e. if the screen contains no structure comparable with the wavelength. Under these conditions, the function $\rho_R(\xi)$ describing the intensity pattern on any plane parallel to the screen is the same at all distances, and it will therefore be representative of the irregularities in the screen itself.

It is important to notice that the form of $|A(s)|^2$ cannot be deduced uniquely from a knowledge of $\rho_R(\xi)$. This is because the upper integral in (3) may be negative for some values of ξ , and it is therefore not permissible to take the square root of each side of the equation and deduce that $|A(s)|^2$ is the Fourier transform of $\sqrt{\rho_R(\xi)}$. This difficulty can be overcome by assuming a standard analytical form for the function $|A(s)|^2$ which is such that a family of curves of varying angular spread are obtained by the variation of a single parameter. The corresponding family of curves for $\rho_R(\xi)$ can then be deduced from (3), and the member of this family which best fits the experimental results can be determined. This method is developed in §§ 3 and 4.

§ 3. A STANDARD FORM OF ANGULAR DISTRIBUTION

The function $\cos^n \theta$ will be used to represent a family of curves of varying angular spread (n being a variable parameter). This function is convenient because various correction factors of the same form appear in the calculations, as shown later. It may readily be shown that for n large (the case which will

* If the quantity inside the large brackets in (3) is denoted by ω , then a more accurate expression for $\rho_R(\xi)$ is $\pi\{\omega^2 + \omega^4/16\}/4(4 - \pi)$ (Burgess and Uhlenbeck, unpublished, see B.R.S.). Equation (3) as given is accurate to within 10% over the range of ω from 0 to 1.

mainly be of interest) the function approaches a Gaussian-distribution of standard deviation $1/\sqrt{n}$. Thus

$$\cos^n \theta \rightarrow \exp(-\frac{1}{2}n\theta^2) \quad (n \text{ large, } -\frac{1}{2}\pi < \theta < +\frac{1}{2}\pi). \quad \dots\dots(5)$$

It follows that $\cos^n 2\theta \simeq \cos^{4n} \theta$ for $-\frac{1}{4}\pi < \theta < +\frac{1}{4}\pi$; this result will be required later.

It is convenient to have a simple measure of the 'spread' of the function $\cos^n \theta$ and for this purpose we will use the angle θ_0 at which the function is equal to $\frac{1}{4}$. (Since the function will be used to represent the angular spread of *power*, this is the angle at which the *amplitude* has fallen to one-half.) The relation between the parameters θ_0 and n is thus $\cos^n \theta_0 = \frac{1}{4}$ and is plotted in Figure 2.

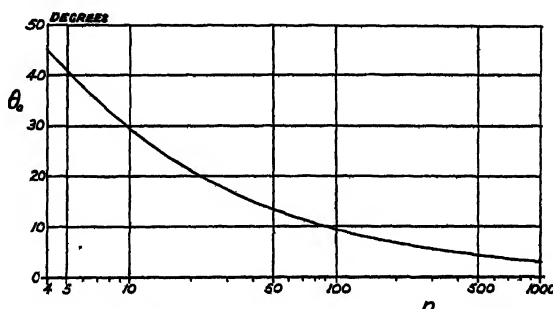


Figure 2. Relation between the two parameters θ_0 and n , as given by the equation $\cos^n \theta_0 = \frac{1}{4}$.

When n is large, and the distribution is approximately Gaussian, we have $\theta_0 = 1.67(1/\sqrt{n})$ i.e. the half-width for amplitude is a constant factor times the standard deviation $1/\sqrt{n}$.

Let the energy received on a detector responding to H_y or E_y and sensitive to the range of angles θ to $\theta + d\theta$ be $W(\theta) d\theta$, where $W(\theta) = \cos^n \theta$. Then since $s = \sin \theta$ so that $ds = \cos \theta d\theta$ we can express this energy as $[A(s)]^2 ds = (1 - s^2)^{(n-1)/2} ds$, and hence, by equation (3),

$$\rho_R(\xi) \propto \left[\int_{-1}^{+1} (1 - s^2)^{(n-1)/2} \cos(2\pi s\xi) ds \right]^2. \quad \dots\dots(6)$$

This gives the auto-correlation function for amplitude as a function of distance over the ground for the assumed form of angular distribution of received power.

The same result will now be obtained by an alternative method. The energy received by a detector sensitive to angles between θ and $\theta + d\theta$ may be regarded as coming from the region $AB = dx$ of the screen in Figure 1(c). Now $d\theta = (\cos \theta dx)/\rho$ and the received power is therefore

$$\frac{W(\theta) \cos \theta dx}{\rho}. \quad \dots\dots(7)$$

This is an expression for the power received from the element dx of the screen. As expected, the power scattered is proportional to the size of the element, and the factor $1/\rho$ may be interpreted as the spreading of the cylindrical wave from the element. The remaining factor depending on angle is the 'scattering polar diagram' which must be attributed to the element.

Thus we may regard the screen as built up of a number of scattering elements, each having a certain scattering polar diagram, and randomly placed in a horizontal plane. The advantage of this approach arises when the screen is two-dimensional and the illumination is non-uniform, but we will first show that it gives a correct result in the simple case of uniform illumination of a one-dimensional screen by a plane wave.

Let the diffracting screen move horizontally with velocity V in the direction of the X -axis. The diffraction pattern observed on any horizontal plane sweeps past a fixed observing point with the same velocity, and the fluctuations with distance are converted into fluctuations with time. Let the auto-correlation function for the fluctuations in time be $\rho_R(\tau)$. This function will be related to the function $\rho_R(\xi)$ by the transformation $\lambda\xi = V\tau$. Let the time fluctuations have a power spectrum $W(f)$. Then we have the result (analogous to equation (3)) that

$$\rho_R(\tau) = \frac{\left[\int_{-\infty}^{+\infty} W(f) \cos(2\pi f\tau) df \right]^2}{\int_{-\infty}^{+\infty} W(f) df} \quad \dots\dots(8)$$

We now proceed to find the form of the function $W(f)$ by a consideration of the Doppler frequency shift of the power scattered from each element of the moving screen. The power scattered from the element dx in Figure 1(c) suffers a frequency shift f given by

$$f = \pm \frac{V}{\lambda} \sin \theta = \pm k s \quad \dots\dots(9)$$

where $k = V/\lambda$. The power received from this element for the particular form of $W(\theta)$ postulated above is, from (7),

$$\cos^n \theta \cos \theta dx / \rho = \cos^{n-1} \theta ds.$$

Hence, from (9) we have

$$W(f) \propto \left(1 - \frac{f^2}{k^2} \right)^{(n-1)/2} df \quad (-k < f < +k)$$

and thus

$$\rho_R(\tau) \propto \left[\int_{-k}^{+k} \left(1 - \frac{f^2}{k^2} \right)^{(n-1)/2} \cos(2\pi f\tau) df \right]^2.$$

To obtain $\rho_R(\xi)$, we put $\tau = \lambda\xi/V$. If, also, we put $f/k = u$ and note that $k = V/\lambda$, the equation gives

$$\rho_R(\xi) \propto \left[\int_{-1}^{+1} (1 - u^2)^{(n-1)/2} \cos(2\pi u\xi) du \right]^2.$$

This result agrees with equation (6) obtained by the direct method.

§ 4. EXTENSION TO A TWO-DIMENSIONAL SCREEN WITH NON-UNIFORM ILLUMINATION

The method of calculation by Doppler shifts will now be extended to the following problem which is much nearer to the one actually encountered in the experiments. Let a random screen be illuminated from a transmitting aerial which has a polar diagram $\cos^2 \theta$ (for power). Let the screen be such that if

illuminated with a *plane* wave, it would scatter power so that the amount received per unit solid angle by a detector varies with θ in a manner given by the function $W(\theta) = \cos^q \theta$. Let the reflected wave be received on an aerial having a polar diagram $\cos^r \theta$ (for power). Then the power received from the element $dx dy$ in Figure 3 (a) may be written

$$dW = \underbrace{\left\{ \frac{\cos^t \theta \cos \theta}{\rho^2} \right\}}_{\text{power flux on the element.}} \underbrace{\{\cos^{4q} \theta \cos \theta dx dy\}}_{\text{scattering power of the element. (cf. equation (7)).}} \underbrace{\{\cos^r \theta\}}_{\text{receiving aerial sensitivity.}} \underbrace{\left\{ \frac{1}{\rho^2} \right\}}_{\text{spherical spreading of the wave from the element.}}$$

The term in this equation which represents the scattering power of the element requires further explanation. Since the elements are illuminated obliquely, equation (7) cannot be used directly. It is not possible to predict rigorously the form of the scattering function from a knowledge of the function for normal illumination, but it will be assumed that the main effect is to turn the polar diagram through an angle θ so that its maximum lies in the direction of the geometrically reflected ray, as shown in Figure 3 (b). The response will therefore

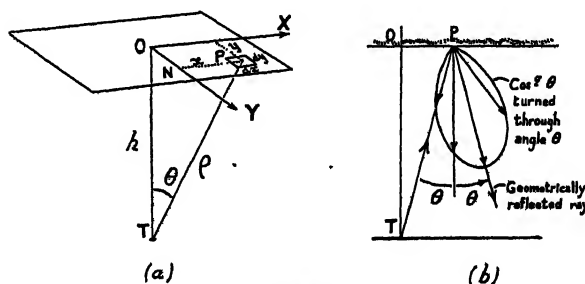


Figure 3.

(a) Diagram to show the notation used for a two-dimensional random screen.

(b) The polar diagram of an element of the screen when illuminated from a transmitter at T.

be given by writing 2θ instead of θ in the term $\cos^q \theta$, or, better, by replacing q by $4q$, which gives a function which is a good approximation to the function $\cos^q 2\theta$ in the range $-\frac{1}{4}\pi < \theta < +\frac{1}{4}\pi$, and which falls smoothly to zero when θ lies outside this range. Thus we obtain the term for the scattering power of the element in the form given above.

The power received from the element $dx dy$ may thus be written

$$dW = \cos^{(t+4q+r+6)} \theta dx dy \quad \dots\dots (10)$$

since $1/\rho^4 \propto \cos^4 \theta$.

Now the solid angle subtended by the element is

$$d\Omega = \frac{\cos \theta dx dy}{\rho^2} \propto \cos^3 \theta dx dy.$$

Thus the power may be written

$$dW \propto \cos^n \theta d\Omega, \quad \dots\dots (11)$$

where

$$n = t + 4q + r + 3. \quad \dots\dots (12)$$

Equation (11) gives the resultant angular distribution of received power. We now wish to relate this to the auto-correlation function of amplitude over the ground, as in the one-dimensional case. It may be noted that as far as the

further development of the present section is concerned, it is possible to take equation (11) as the starting point, and thus avoid the approximations used above. Later, in § 6, however, we wish to deduce the value of q from the observed value of n , and the approximations cannot then be avoided.

The analysis will now be developed as in § 3. Let the diffracting screen move horizontally with velocity V in the direction of the X -axis. The diffraction pattern moves with velocity $2V$ for the mode of illumination here considered, so that the transformation from space to time auto-correlation functions is made by writing $\lambda\xi = 2V\tau$ in this case.

The energy returned from the element at P in Figure 3 (a) will have a Doppler frequency shift given by

$$f = \frac{2V}{\lambda} \cos \widehat{NPT} = \frac{kx}{(x^2 + y^2 + h^2)^{1/2}},$$

where $k = 2V/\lambda$. Thus elements of the screen which produce a constant frequency shift f lie on the hyperbola

$$\frac{x^2}{f^2} = \frac{x^2 + y^2 + h^2}{k^2}. \quad \dots\dots(13)$$

This hyperbola is shown in Figure 4. Consider a strip of the screen of width dy

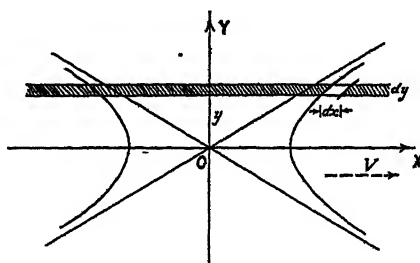


Figure 4. Hyperbola of constant Doppler frequency shift.

and distant y from the X -axis. From (13) we have

$$dx = \frac{(x^2 + y^2 + h^2)^{3/2}}{k(y^2 + h^2)} df.$$

This equation gives the range dx within the strip corresponding to the frequency interval df . The power received from the element $dx dy$ is from (10) and (12)

$$dW \propto \frac{dx dy}{(x^2 + y^2 + h^2)^{(n+3)/2}}.$$

The contribution to the power spectrum for the frequency range f to $f + df$ is therefore

$$dW(f) df \propto \frac{dy df}{(x^2 + y^2 + h^2)^{n/2} (y^2 + h^2)},$$

where (x, y) lies on the hyperbola appropriate to the frequency f , i.e. (x, y) satisfies equation (13). Using this fact we have

$$dW(f) \propto \frac{dy}{k(y^2 + h^2)^{(n+2)/2}} \left(1 - \frac{f^2}{k^2}\right)^{n/2}.$$

Thus

$$W(f) \propto \left(1 - \frac{f^2}{k^2}\right)^{n/2} \int_{-\infty}^{+\infty} \frac{dy}{(y^2 + h^2)^{(n+2)/2}}.$$

The integral in this equation is a numerical factor of no interest. Thus

$$W(f) \propto \left(1 - \frac{f^2}{k^2}\right)^{n/2}. \quad \dots\dots(14)$$

This expression holds for $-k < f < k$, and $W(f)$ is zero for f outside these limits. From (8)

$$\rho_R(\tau) \propto \left[\int_{-k}^{+k} \left(1 - \frac{f^2}{k^2}\right)^{n/2} \cos(2\pi f\tau) df \right]^2$$

and, writing $f/k = u$, $\lambda\xi = 2V\tau$ and $k = 2V/\lambda$, we have

$$\rho_R(\xi) \propto \left[\int_{-1}^{+1} (1 - u^2)^{n/2} \cos(2\pi u\xi) du \right]^2. \quad \dots\dots(15)$$

This expression gives the required auto-correlation function. It has been evaluated for small values of n by means of a reduction formula, and for large values of n by an approximate method. The resulting family of curves is shown in Figure 5.

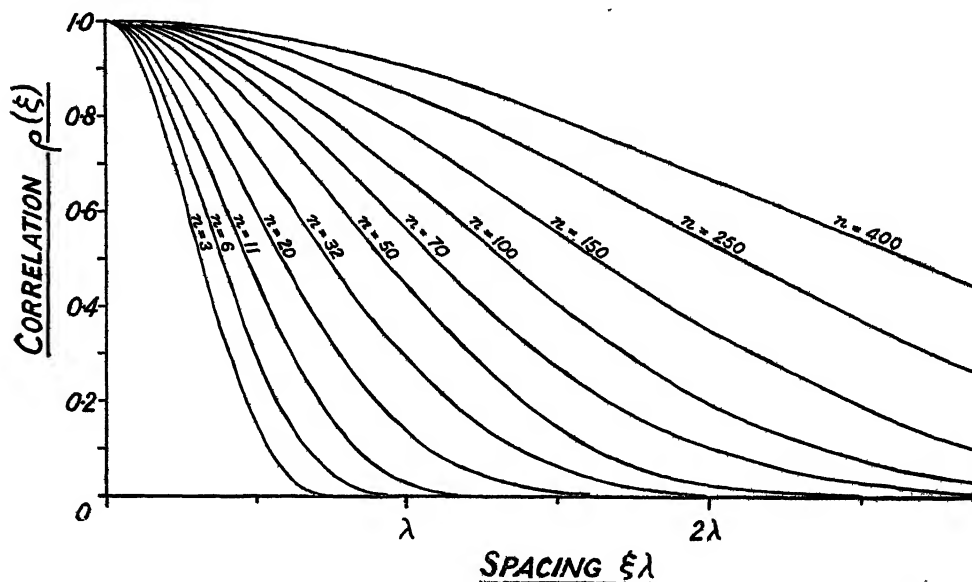


Figure 5. Theoretical auto-correlation functions $\rho_R(\xi)$ assuming the distribution of received power with angle to be given by the function $W(\theta) = \cos^n \theta$.

In practice the auto-correlation function $\rho_R(\xi)$ is inconvenient for computations, because it is based upon the average value of the product of $R(x)$ and $R(x+\xi)$ (equation (4)). It is more convenient to use a measure of correlation $\Delta(\xi)$ based on the average value of the *difference* of the amplitudes at the two points. An auto-correlation function of this type has been used in the study of random noise (Fürth and MacDonald 1947). We define

$$\Delta(\xi) = |R(x) - R(x+\xi)|.$$

When it is necessary to refer to $\Delta(\xi)$ in words, we will call it the 'average difference' for brevity; strictly, it is the average value of the modulus of the difference.

It can be shown (McNicol 1949), subject to certain conditions which are usually satisfied in practice, that $\Delta(\xi)$ and $\rho_R(\xi)$ are related by the equation

$$\Delta(\xi) = 0.59\bar{R}[1 - \rho_R(\xi)]^{1/2}. \quad \dots\dots(16)$$

By means of this equation, curves of $\Delta(\xi)/\bar{R}$ have been plotted, and are shown

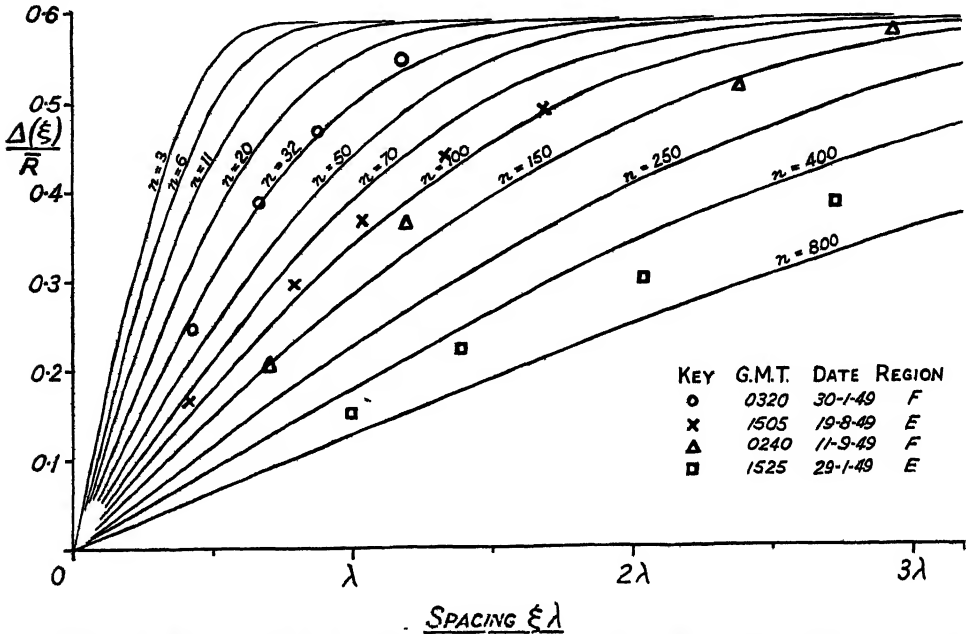


Figure 6. Curves of the difference-type correlation function $\Delta(\xi)$ corresponding to the product-type functions $\rho_R(\xi)$ of Figure 5. Four sets of experimental points, obtained by the method explained in the text, are shown, for comparison with the theoretical curves.

in Figure 6. By use of the graph in Figure 2 it is now possible to derive the curves which relate the angular spread of received power measured by the parameter θ_0 to the average difference of signal amplitude at two observing points at any fixed distance apart. In Figure 7 such curves are plotted for various receiver spacings. These graphs present the results in the most convenient form for comparison with experiment provided it is not necessary to correct for the polar diagrams of the aerials. It appears from Figure 7 that a receiver spacing of one wavelength is about the optimum in the sense that it gives the maximum sensitivity over the range of angles which is mainly of interest, and this spacing has been used most frequently in the experiments.

§ 5. PHASE DIFFERENCES AND RELATION TO DIRECTION FINDING EXPERIMENTS

The average difference of phase of the signals observed at a pair of spaced receivers will now be calculated. MacDonald (1949) has considered the analogous problem for a random time variable, and has shown that, if the fluctuation has a power spectrum of Gaussian form, with standard deviation σ , then

$$|\phi(t) - \phi(t + \tau)| = 2\pi\sigma\tau \quad \text{for } \tau \text{ small,}$$

where $\phi(t)$ is the instantaneous phase observed at time t . A comparison of equations (3) and (8) shows that the analogous equation for the fluctuation of phase over the ground is

$$|\phi(x) - \phi(x + \xi)| = 2\pi\sigma\xi$$

provided that $|A(s)|^2$ is Gaussian, with standard deviation σ . With the form for $|A(s)|^2$ assumed in § 3, this is the case for n large and $\sigma \simeq 1/\sqrt{n} = \theta_0/1.67$. Thus

$$|\phi(x) - \phi(x + \xi)| = 2\pi\theta_0\xi/1.67. \quad \dots\dots(17)$$

This equation, then, gives the required average phase difference for a pair of receivers a distance ξ wavelengths apart, in terms of the parameter θ_0 specifying the spread of the angular spectrum.

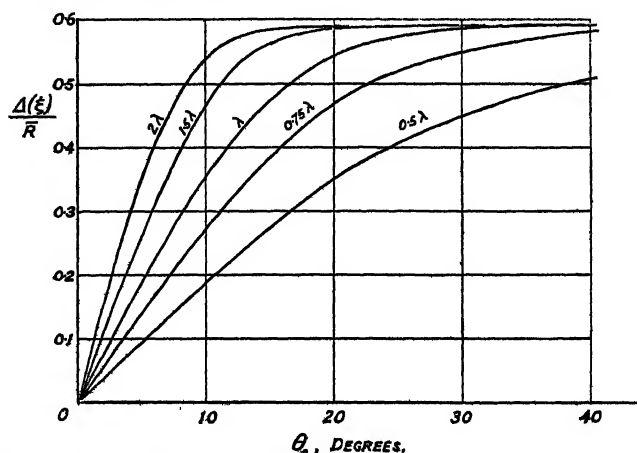


Figure 7. Curves showing the average difference of signal amplitude at a pair of receivers at various distances apart in terms of the parameter θ_0 specifying the degree of angular spreading of the downcoming wave. By means of these curves θ_0 can be obtained directly from the fading records at two such receivers.

In a direction-finding experiment, the instantaneous phase difference between the receivers would be interpreted as an instantaneous direction of arrival of the wave θ_i . Provided that the angles of deviation are small, θ_i is related to the observed phase differences by the equation

$$\phi(x) - \phi(x + \xi) = \frac{2\pi}{\lambda} \lambda \xi \theta_i.$$

Thus, from (17), we have

$$2\pi\xi|\theta_i| = 2\pi\theta_0\xi/1.67$$

or

$$|\theta_i| = 0.6\theta_0. \quad \dots\dots(18)$$

This equation relates the *mean* bearing deviation (without regard to sign) to the half-width θ_0 of the angular spectrum. The extreme deviations observed, will, of course, be much greater than θ_0 .

§ 6. CORRECTION FOR AERIAL POLAR DIAGRAMS AND DEDUCTION OF THE SIZE OF THE IRREGULARITIES

We have so far assumed the angular distribution of received power to be given by the function $\cos^2\theta$, and have considered the amplitude and phase differences at spaced receiving points which arise from this distribution. As a result it was possible to construct curves from which the parameter n or the

closely related parameter θ_0 may be deduced from observations. That is to say n may be found by observing which curve in Figure 6 comes nearest to the point representing the observed correlation for a known receiver spacing, or which best fits a complete experimental correlation curve if this is available.

If however we want to go further than the determination of the actual received angular power density, and wish to express the amount of scattering by the ionosphere in a form independent of aerial properties, it is necessary to consider the separate contributions to the value of n as expressed by equation (12).

In the first place we can correct for the effect of aerial polar diagrams. This is only an important correction when the spreading of the downcoming wave is relatively large. The corrected value $n' = n - t - r$ would be obtained in an experiment using isotropic aerials for which $t = r = 0$. The values of θ_0 given in the next section are those corresponding to this corrected value of n . The experimental arrangement usually employed a horizontal dipole for the transmitter and loop aerials for the receivers. A dipole in free space has a $\cos^2 \theta$ polar diagram (for power) in a plane containing its own length and a circular diagram in the plane at right angles to its length; an average of $\cos \theta$ may be assumed. Owing to the presence of the ground an additional $\cos^2 \theta$ is superimposed in both planes and we therefore take $\cos^3 \theta$, or $t = 3$. A loop aerial (magnetic dipole) will behave on the ground as if in free space, and we may take an average of $\cos \theta$, or $r = 1$.

To correct the results to the case of plane wave illumination it is only necessary to evaluate the value of q from equation (12). Thus

$$q = \frac{n' - 3}{4} = \frac{n - t - r - 3}{4}$$

and the function $\cos^2 \theta$ gives by the definition of q the angular distribution of received power which would be obtained if the ionosphere had plane waves incident upon it, and isotropic aerials were employed. This last step has taken into account the different range and obliquity factors involved in this case.

It is now possible to deduce something about the size of the ionospheric irregularities because we know that for plane wave illumination the autocorrelation function of the diffraction pattern on the ground is the same as that of the irregularities themselves, provided that no evanescent waves are present. We therefore define a characteristic length or 'size', l , of the irregularities as the distance between two points on the ground for which the correlation is 0.5, in the idealized case of plane wave illumination and 'all-round' aerials. (This is similar to a definition often used for the size of eddies in the theory of turbulence.) The value of l can be found from the curve in Figure 5 corresponding to the deduced value of q in place of n , by noting the spacing at which $\rho = 0.5$.

The main correction which involves a factor of four in the power of $\cos \theta$ is equivalent to assuming that the irregularities in the diffraction pattern actually produced are twice as large as those in the diffracting screen itself, because the illumination is from a point source at the ground rather than by a plane wave.

§7. EXPERIMENTAL RESULTS

The apparatus which has been used to study the fading at spaced receivers is in principle the same as that described by Mitra (1949). (Some improvements in experimental technique will be described elsewhere.) Three receivers are

placed at the corners of a right-angled triangle, so that one pair lies in a N-S line and another pair in an E-W line, the spacing of each pair being approximately 130 metres. The fading of a single echo from a local pulse transmitter is recorded photographically at each point, the three records being recorded on a single film for ease of comparison.

It is first desirable to confirm that the assumed form of angular distribution provides a good approximation to those which occur in practice. This can be done by comparing the detailed shape of the theoretical auto-correlation functions $\Delta(\xi)$ of Figure 6 with experimentally determined curves. It would appear at first sight that in order to determine the latter, it would be necessary to have a large number of receivers at various spacings along a line, so that the correlation could be determined for a large number of spacings. Fortunately this is not so, for the following reason. If three receivers are placed at the corners of a right-angled triangle, it is possible by a comparison of the three fading records to detect any uniform movement of the diffraction pattern over the ground, and also to determine whether the pattern moves without change of form, or whether there are random changes present in addition to the steady movement (Briggs, Phillips and Shinn 1950). It is found that there are occasions when the random changes are negligible compared with those due to the steady drift. The fading observed at a fixed point is then due to an unchanging diffraction pattern moving over the ground, and the auto-correlation function for the fading in time, $\Delta(\tau)$, will be the same as the auto-correlation function for distance, $\Delta(\xi)$, subject to the transformation $\lambda\xi = V\tau$, where V is the drift velocity of the diffraction pattern. The velocity V can be determined in the manner described by Briggs, Phillips and Shinn. Thus, in order to obtain the desired check on the assumed angular distribution function, it is only necessary to evaluate $\Delta(\tau)$ on an occasion when the fading has been shown, from the three receiver experiment, to be due predominantly to a drifting diffraction pattern.

The method which has been used to compute $\Delta(\tau)$ in practice is as follows. A tracing of the fading curve is made and is placed over the original fading record with a time shift equal to the desired value of τ . The included area between the original and the shifted curve is measured with a planimeter. $\Delta(\tau)$ is equal to this area divided by the length of the record. The area underneath one fading record is also measured, and \bar{R} is obtained by dividing this area by the length of the record. $\Delta(\tau)/\bar{R}$ is then plotted against τ . Some of the experimental points are shown in Figure 6 for comparison with the theoretical curves. It will be seen that the value of n can be determined for each example with some accuracy.

This method is restricted to occasions when the fading is due entirely to a movement of an unchanging diffraction pattern, and is also too laborious for general application. In general only a single point on the $\Delta(\xi)$ curve is known, obtained from the average difference between the fading records at the two receiving points. We have measured this quantity for a large number of records with a planimeter, and as a result of the work it has been found that an experienced observer is able to estimate the average difference between the records (without any measurement at all) to an accuracy of about 20%. For compiling statistical results this accuracy is sufficient, and in view of the very large number of records which have been obtained, it has been necessary to use this method for the largest proportion of the records, which were obtained from January 1949–January 1950 on a frequency of 2.4 Mc/s. Recently we have developed a method of recording the

required average difference directly, and this method has been used to obtain further results on a frequency of 4.8 Mc/s. (January–March 1950).

From Figure 7 each record yields a value of θ_0 , the angle to the half-amplitude point of the function representing the cone of received rays. In Figure 8 histograms* are given to show the number of occasions on which different values of θ_0 were observed for regions F, E, and E_s . A value of θ_0 can be obtained from the E–W pair of receivers and from the N–S receivers, and on the average the two values are found to be equal, though on individual occasions significantly different values may be obtained. In constructing the histograms the two values have been averaged. It will be seen that for 2.4 Mc/s., regions E and F

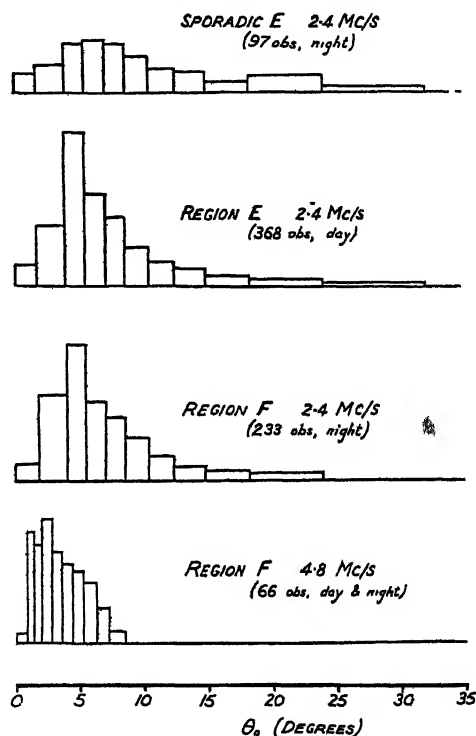


Figure 8. Histograms showing the frequency of occurrence of different values of the parameter θ_0 which specifies the degree of angular spreading of the downcoming wave. Separate histograms are shown for regions E_s , E, and F (2.4 Mc/s.), and region F (4.8 Mc/s.).

give similar histograms with a value of θ_0 of about 5° most frequently observed. The histogram for region E_s has a less pronounced peak and also extends to larger values of θ_0 . On this frequency the results on region E are entirely day-time observations, and the results on region F and region E_s are entirely night-time. On the frequency of 4.8 Mc/s., region F has been observed by both day and night. As predicted by the theory, the angular spread is smaller, and the peak in the histogram occurs at a value of θ_0 of about 2.5° . The number of occasions on which echoes from region E_s were observed on this frequency was too small to yield statistical results.

* Owing to the various ways in which the original data have been transformed and corrected, the histograms given in this paper have unequal intervals of the abscissae. In every case, the number of observations contained between two values of the abscissae in the final histogram has been divided by the length of the interval, so that the ordinates of the histograms are proportional to the number of observations per unit range of the abscissa.

The results shown in Figure 8 relate entirely to the properties of the cone of rays falling on the aerial system, and each histogram may be expected to hold only for the frequency on which it was obtained. If histograms are constructed for the sizes of the ionospheric irregularities deduced by the method described in § 6, these should hold for all frequencies, except in so far as irregularities at different heights in the ionosphere with possibly different statistical properties may become effective as the frequency is changed.

The histograms of Figure 9 were obtained for a number of occasions on which irregularities of different sizes were observed. The results for the two frequencies 2.4 Mc/s. and 4.8 Mc/s. are seen to be in agreement.

When the fading records at the spaced receivers are highly correlated, it should be pointed out that the procedure used in deducing the size of the irregularities involves a considerable extrapolation of the data, and is dependent entirely on

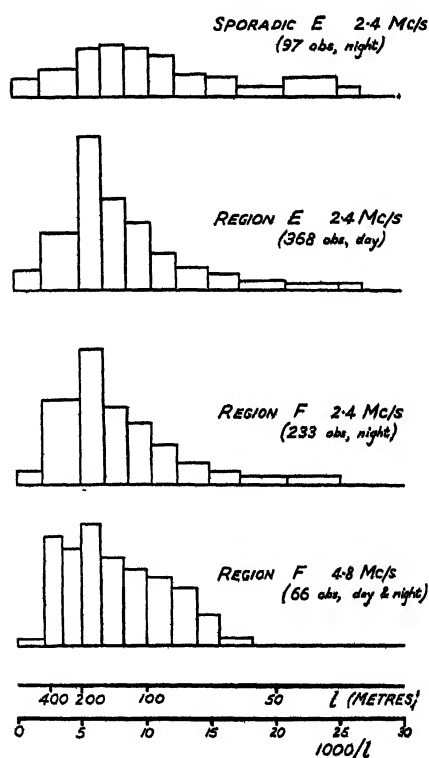


Figure 9. Histograms showing the frequency of occurrence of different values of the characteristic length l which specifies the size of the ionospheric irregularities. Separate histograms are shown for regions E_s, E, and F (2.4 Mc/s.) and region F (4.8 Mc/s.).

the particular shape postulated for the correlation functions. In essence, we are attempting to predict from the small measured difference between a pair of closely spaced receivers at what distance apart the receivers would have to be placed in order to obtain records almost uncorrelated. No high accuracy can therefore be claimed, though we have already given some indirect evidence which suggests that the assumed shape of the correlation functions is, in fact, obeyed up to quite large distances.

Some implications of the results may be noted. As the most frequent size of the irregularities is found to be of the order of 200 m. it follows that on radio wavelengths such that $\lambda > 200$ m. the ionosphere will behave as a 'completely rough' screen, as defined by B.R.S.

With the notation of §4, such a screen would have $q=0$, and if 'all-round' aerials were used, would have $n=3$. The auto-correlation function would be given by the curve marked $n=3$ in Figure 5. The fading on these wavelengths should, therefore, be practically uncorrelated at a receiver spacing of half a wavelength. For radio wavelengths such that $\lambda < 200$ m., however, the auto-correlation function becomes independent of wavelength, and therefore the correlation at one half wavelength spacing becomes greater as the wavelength decreases. In this range, receivers at a fixed absolute spacing tend to show the same correlation, independent of the wavelength. From the results of experiments with three receivers, interpreted as described by Briggs, Phillips and Shinn (1950), we believe that the predominant cause of fading in the short-wave

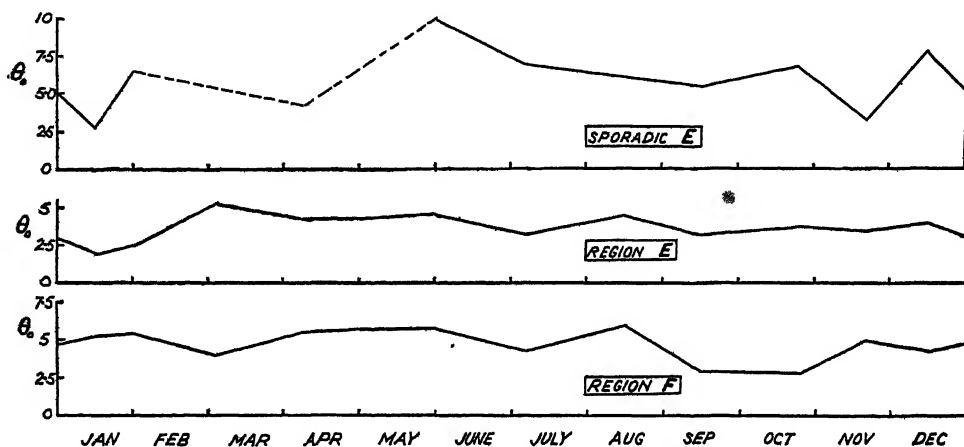


Figure 10. Average values of the parameter θ_0 observed on three days of each month through the year (1949-50), for a frequency of 2.4 Mc/s. No significant seasonal variation is apparent. The results for region E relate to day-time observations, and the results for regions F and E_s to night-time observations.

range is a tendency for the diffraction pattern to drift over the ground due to the presence of an ionospheric wind, and that any random changes are only of secondary importance. If it is assumed that the same mechanism holds also in the range $\lambda > 200$ m., and if we interpret the above results in terms of speed of fading, we see that for $\lambda > 200$ m. we should expect the fading speed to be inversely proportional to wavelength, and for $\lambda < 200$ m. the fading speed should be independent of wavelength. The precise wavelength at which the change-over occurs, may, of course, be different on different occasions. These consequences are at present being tested by more direct experiments. As vertical incidence has been used in the experiments the conclusions of this paragraph will not in general apply to oblique incidence.

It is of interest to know whether the measurements show any important diurnal or seasonal variations. Results extending over a year are available for a frequency of 2.4 Mc/s. Figure 10 shows the average value of the parameter θ_0 obtained for each month of the year 1949-1950, for the three regions, E, F, and E_s.

No significant seasonal variation is apparent, nor has any significant diurnal variation been found. The diurnal variation on this frequency is complicated by the changing time at which the transition from region E to region F occurs at different times of the year. However, it can be stated that any diurnal variation is much smaller than the day to day and hour to hour variability.

Two limitations to the present theory should be noted: (i) When the scattering is very extreme, the received pulse is appreciably spread in range and no unique value can be assigned to its amplitude. This effect occurs mainly near the penetration frequencies of the regions, and also occasionally during the later part of the night when the 'spread F' phenomenon occurs. (ii) If the fading is 'shallow' (displaced Gaussian rather than Rayleigh amplitude distribution) there must be an appreciable specular component (McNicol 1949). This corresponds to a line in the angular spectrum at $\theta = 0$. The assumed form of angular spectrum then breaks down and the results given in the present paper are inapplicable. We have rejected a small proportion of records which come under one or other of these two headings.

ACKNOWLEDGMENTS

This work forms part of a programme of Radio Research at the Cavendish Laboratory supported by a grant from the Department of Scientific and Industrial Research. We are also in receipt of individual maintenance grants from the same Department.

REFERENCES

- BOOKER, H. G., RATCLIFFE, J. A., and SHINN, D. H., 1950, *Phil. Trans. Roy. Soc. A*, **262**, 579.
BRIGGS, B. H., PHILLIPS, G. J., and SHINN, D. H., 1950, *Proc. Phys. Soc. B*, **63**, 106.
FÜRTH, R., and MACDONALD, D. K. C., 1947, *Proc. Phys. Soc.*, **59**, 388.
MACDONALD, D. K. C., 1949, *Proc. Camb. Phil. Soc.*, **45**, 368.
MCNICOL, R. W. E., 1949, *Proc. Instn. Elect. Engrs.*, Pt. III, **96**, 517.
MITRA, S. N., 1949, *Proc. Instn. Elect. Engrs.*, Pt. III, **96**, 441.
PAWSEY, J. L., 1935, *Proc. Camb. Phil. Soc.*, **31**, 125.
RATCLIFFE, J. A., and PAWSEY, J. L., 1933, *Proc. Camb. Phil. Soc.*, **29**, 301.

Periodic Fading of Short-Wave Radio Signals

By S. R. KHAISTGIR AND P. M. DAS

Benares Hindu University, Benares, India

MS. received 23rd September 1949, and in final form 26th April 1950

ABSTRACT. Periodic fading patterns were recorded photographically with Calcutta signals of frequency 4,840 kc/s. received at Dacca (distance 240 km.) during the evening and early night hours of December 1948 and January 1949. The main features in the experimental conditions were: (i) the operating frequency was much less than the maximum usable frequency (M.U.F.) for the F layer transmission, (ii) the frequency was slightly greater than the M.U.F. for the ordinary wave transmission through the E layer and (iii) it was slightly less than the M.U.F. for the extraordinary wave transmission through the E layer between the transmitting and receiving stations. The following patterns of periodic or rhythmic fading were observed:

(i) Sinuous fading of comparatively quick period: this is considered to be of magneto-ionic origin, due to the interference between the upper and lower trajectory extraordinary waves in the E layer, the ordinary waves having passed through the E layer.

(ii) Periodic or rhythmic fading of comparatively slow period: the slow periodic fading is considered to be due to the beat-effect between the singly and doubly reflected waves from the F2 region or between the singly reflected waves from the E and F2 regions, the two interfering waves in different directions having suffered different amounts of Doppler change of frequency due to the vertical movement of the ionospheric layer or layers. The vertical velocity of the ionosphere as computed from this view agrees with the observed value.

(iii) Slow periodic fading with superposed ripples: this was observed when the ionospheric conditions were favourable for the simultaneous occurrence of the magneto-ionic type of sinuous fading and the Doppler beat type of slow periodic fading. In a few patterns of periodic fading there was evidence of extremely high frequencies (4–12 cycles/second) the origin of which is unknown.

§ 1. INTRODUCTION

THE work of Appleton and Beynon (1947) has shown that the interference of the ordinary and extraordinary components of a short-wave signal caused by magneto-ionic splitting in the ionosphere gives rise to rhythmic fading at a time, usually during the evening or early morning hours, when there is a steady decrease or increase of electron density in the ionosphere. For this type of rhythmic fading the ionospheric ionization should be just enough for a single reflection to take place between the transmitting and receiving stations; in other words the frequency of the wave should be in the neighbourhood of the maximum usable frequency (M.U.F.) for the relevant ionospheric layer. Rhythmic fading of a different origin has also been reported by Banerjee and Mukerjee (1948) and Banerjee and Singh (1948, 1949) under conditions of high ionospheric ionization. In the present paper an account is given of an investigation on rhythmic or periodic fading which throws some light on the origin of such fading observed with short-wave signals. The investigation was carried out with Calcutta signals of frequency 4,840 kc/s. received at Dacca at a distance of 240 km. from Calcutta in the evening and early night hours of December 1948 and January 1949.

The observed rhythmic or periodic fading patterns were recorded photographically. The paper gives a description of the experimental method and a discussion of the experimental results. In interpreting the results a new viewpoint has been advanced which explains the rhythmic fading observed under certain conditions, when the Appleton-Beynon type of periodic fading was not

§ 2. EXPERIMENTAL ARRANGEMENT

(i) Receiving Set

A T.R.F. receiver having one stage of R.F. amplification followed by a grid-leak detector and an A.F. power amplifier (to work a loudspeaker for facilitating tuning) was used with a big frame aerial. The receiver and the aerial tuning condenser were inside a shielded box. A suitable mirror galvanometer kept outside the shielded box was inserted in the anode circuit of the detector valve with a conventional arrangement for balancing out the no-signal anode current.

(ii) Recording Apparatus

The recording arrangement was similar to that adopted by Subba Rao and Subramanyam (1942) in their work on atmospherics.

The general disposition of the recording apparatus is shown in Figure 1. With the help of the condensor lenses L_1 and L_2 placed in the paths of the incident and reflected beams to and from the galvanometer mirror M, the light from a point-o-lite was converged to a sharp focus on a bromide paper fixed round the rotating

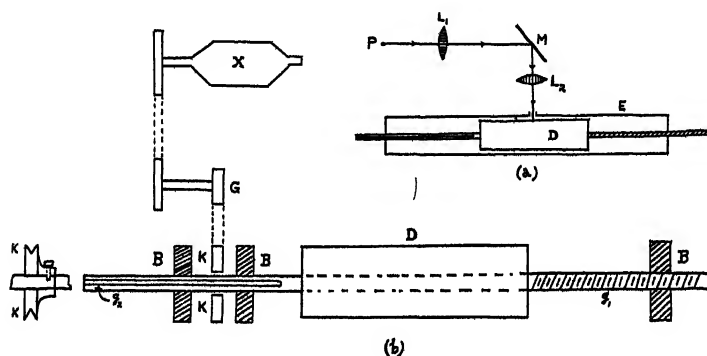


Figure 1 (a). General disposition of apparatus:

P—Point-o-lite, M—Galvanometer mirror, L_1 , L_2 —Convex lenses,
D—Rotating Drum, E—Light-tight enclosure.

Figure 1 (b). Working of the drum:

X—Electric motor, G—Reduction gear, B—Bearings, g_1 —Spiral groove,
 g_2 —Linear groove, KK—Pulley fitted with a guide pin.

drum D. The drum, which was made of wood, was cylindrical in shape with a circumference of 12 in. and length 10 in. It was enclosed inside a long light-tight box E provided with an opening for admitting light for the photographic recording.

The brass axle of the drum, which projected on either side through a length of 12 in., was supported on suitable bearings. Along the axle-rod on one side there was a spiral groove of pitch 0.9 cm. and on the other side a linear groove 3 mm. deep. The bearing which supported the axle-rod with the spiral groove on it had an attached pin fitting into the spiral groove. Round the axle-rod with the linear groove on it there was a pulley KK with a guide pin which fitted into the linear groove and could move along it. As the pulley was made to rotate slowly with the help of an electric motor and the reduction gear, the guide pin on the pulley fitting into the linear groove caused the axle and the drum to rotate at the same time. Besides this rotation there was also a linear motion of the drum along its axis due to the spiral groove and the attached bearing. The linear movement was maintained as the guide-pin on the pulley moved along the linear groove.

This arrangement enabled the drum to move laterally through a distance of 0.9 cm. for one complete rotation. Hence the pattern marked out by the galvanometer spot on the bromide paper, when undeviated by the signals, consisted of slightly inclined parallel lines along the length of the paper.

The speed of rotation of the drum was such as to furnish a time scale of 15–40 cm. per minute. For each run of the recording drum, the speed was determined very accurately.

§ 3. EXPERIMENTAL RESULTS

With Calcutta signals of frequency 4,840 kc/s. received at Dacca during the evening and early night hours, the following rhythmic fading patterns were observed :

(i) *Sinusous Fading of Comparatively Quick Period*

Some typical sinuous fading patterns obtained with carrier waves and with fixed tone modulation are shown in Figure 2*. Some records of sinuous patterns which were observed with the usual modulated waves during the time of a broadcast programme are illustrated in Figure 3. The frequency of these patterns ranged roughly between 25 and 60 cycles/minute. The sinuous patterns were usually observed in the early hours of the evening.

(ii) *Periodic Fading of Comparatively Slow Period*

Some representative patterns of this type are shown in Figure 4. The *quasi*-frequency of this type ranged roughly between 2 and 10 cycles/minute. With the speed of the recording apparatus, still lower frequencies were not easily discernible.

(iii) *Slow Periodic Fading with Superposed Secondary Structure*

The secondary structure superposed on the slow periodic pattern was sinuous or periodic and occasionally random. A few patterns of this type are shown in Figure 5. The *quasi*-frequency of the primary pattern was of the same order as that of the slow periodic fading described under (ii). The frequency of the ripples superposed on the slow pattern was of the same order as that of the sinuous fading described under (i). In a few cases the superposed ripples were as rapid as 80–100 cycles/minute (Figure 5).

The slow periodic fading patterns with or without superposed ripples were usually observed in the later hours of the evening.

In some records the superposed ripples were found to have maximum amplitude at each peak of the primary pattern with a gradually decreasing amplitude on either side of the peak. Such secondary structure can be seen in Figure 5(b). Reference should also be made to extremely high frequencies (4 to 12 cycles/second) observed in a few of the sinuous and other patterns (see Figure 3). This may perhaps be identified with the, 'flutter' phenomena recently reported by Subba Rao and Somāyajulu (1949).

That there was little or no *selective* fading was evident from the fact that the carrier waves with no modulation, with fixed modulation and with the usual programme-modulation yielded more or less similar types of fading patterns. Perfectly sinuous patterns were found mostly with no modulation and with constant modulation (Figure 2). Good sine curves were also observed when the broadcast programme was on (Figure 3). In some records there was evidence of slight distortion in the sinuous patterns obtained when the programme was on.

* For Figures 2–5 see Plates.

It is worthy of note that *pure* random fading with 4,840 kc/s. signals was less frequently observed. The random variation of signal intensity is not, however, the subject matter of the present paper.

The frequencies or *quasi*-frequencies and other details of the different rhythmic patterns illustrated in the paper are given, figure by figure, in the Table.

Fig.	Date	Nature of signals	Time			Frequency (cycles/min.)		Remarks
			I.S.T.			Quick period	Slow period	
			hr.	min.	sec.			
2	16.12.48	Carrier	18	34	20	29.7	—	Sinuous
		Tone	18	35	20	29.7	—	Sinuous
		End of line 5	18	38	40			
3	30.12.48	Programme	18	43		24.5	—	Sinuous
	26.12.48	Programme	19	50		35.4	—	Sinuous
4	26.12.48	Programme	19	37		—	9.6	Slow periodic
5	19.12.48	Programme	18	49		48.8	5.0	Slow periodic with ripples
				to		97.8		
						34.9		
5	18.12.48	Programme	20	12	25	79.2	4.7	Slow periodic with ripples
			Finish	20	36	25		

§4. ORIGIN OF THE OBSERVED RHYTHMIC FADING

The sinuous fading pattern of quick period and the slow periodic fading observed with 4,840 kc/s. signals constituted the two main types of rhythmic fading, the origins of which appear to be quite distinct. The origin of the two main types of rhythmic fading observed in the investigation is discussed below.

(i) *Sinuous or Periodic Fading of Magneto-ionic Origin*

The ionospheric data for Calcutta show that between Calcutta and Dacca, the M.U.F. values for the E and F layer transmission of the ordinary waves were about 4 Mc/s. and 13 Mc/s. during the evening hours of the observation months. The ordinary component of the signal of 4.84 Mc/s. therefore invariably passed through the E layer. The extraordinary components were, however, returned from the E layer in the early evening hours when the electron density of the layer was sufficient for the purpose. Since the frequency of the signal was only slightly less than the M.U.F. value for the extraordinary component in the E layer, the condition was favourable for interference to take place between the lower and upper trajectory extraordinary waves in that layer, giving rise to a periodic type of fading, with periodicity depending upon the difference between the signal frequency and the M.U.F. for the extraordinary wave in the E layer. The observed sinuous patterns of 25-60 cycles/minute were probably caused by interference of this type. It is to be noted that since the signal frequency was *considerably* less than the M.U.F. value for the F layer, the interference between the upper and lower trajectory waves (either ordinary or extraordinary) in that layer was out of the question. The interference between the lower ordinary and the lower extraordinary components in the F layer would not also be able to produce a discernible periodicity.

(ii) *Periodic Fading Due to Doppler Effect in Moving Ionosphere**

When the E layer was 'patchy' in the later parts of the evening it was at times possible with 4,840 kc/s. signals to receive the reflected waves from the F region as well as the extraordinary waves from the E layer. When the electron density of the

* A brief report of this view-point has been published by one of us (Khestrin 1949)

E layer was too low even for the extraordinary component to get returned, the reception would be due to the F layer transmission only. The electron density* of the F layer in the evenings was usually sufficient for both single and double reflection to take place at the F layer (with 4,840 kc/s. signals the minimum electron densities of the F2 region for single and double reflections should be about 2.1×10^5 and 2.8×10^5 electrons/cm³ respectively). Thus in the later parts of the evening the downcoming waves usually consisted of the singly and doubly reflected waves from the F2 region.

With a vertical movement of the ionospheric layers in the evening (or early morning) when there is simultaneous reception of the singly and doubly reflected waves from the F2 region, it is evident that they will suffer different amounts of Doppler change of frequency as they proceed towards the receiving point from distinctly different directions. Thus there will be a difference in the frequencies of the singly and doubly reflected waves from the same moving layer. A similar difference in frequency is also expected in the case of simultaneous single reflections from E and F regions. In either case, the two interfering waves of slightly different frequencies would give a resultant beat-note with a progressively increasing amplitude followed by a corresponding decrease in amplitude of periodic character. When the beat-note is received by the receiver, the output after rectification in the detector stage would constitute the envelope of such resultant beat-note with one side wiped out. This would be similar in appearance to a slow rhythmic fading, the periodicity of which would correspond to the difference in the frequencies of the downcoming waves as determined by the Doppler formula.

Let us first consider the case when the singly and doubly reflected waves from the same layer proceed towards the receiving point. It is to be noted that a singly reflected wave suffers a Doppler change of frequency once and a doubly reflected wave twice. If the angles of incidence corresponding to the singly and doubly reflected waves from the same layer moving with velocity v in the vertical direction are denoted by θ_1 and θ_2 respectively, then the difference in frequency Δf between the two interfering waves is given by

$$\frac{\Delta f}{f} = \frac{2v}{c} (2 \cos \theta_2 - \cos \theta_1)$$

when f is the signal frequency and c the velocity of light. The *quasi*-frequency n of the periodic pattern which is the same as Δf is then obtained from

$$n = \frac{2v}{\lambda} (2 \cos \theta_2 - \cos \theta_1) \quad \dots\dots (1)$$

where λ is the wavelength corresponding to the frequency f . Similarly, in the case of singly reflected waves from the two layers moving with the same velocity v in the vertical direction, the *quasi*-frequency of the periodic fading pattern due to the Doppler beat effect is given by

$$n = \frac{2v}{\lambda} (\cos \phi_1 - \cos \theta_1) \quad \dots\dots (2)$$

* The ionospheric data from the Calcutta observations under Prof. S. K. Mitra show that on the observation days and time, the electron density of the F2 region ranged from 1.55×10^6 to 2.4×10^6 electrons/cm³.

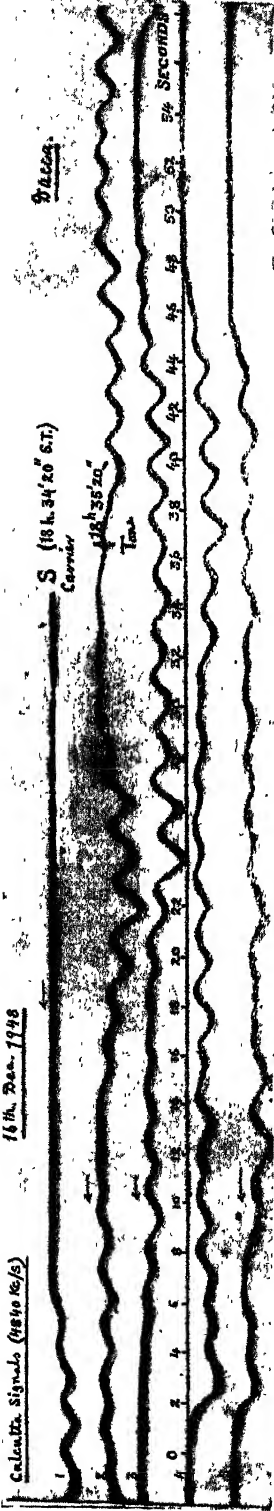


Figure 2. Sinuous patterns. S—Record starts with carrier waves at 18h. 34m. 20s. and tone begins at 18h. 35m. 20s.

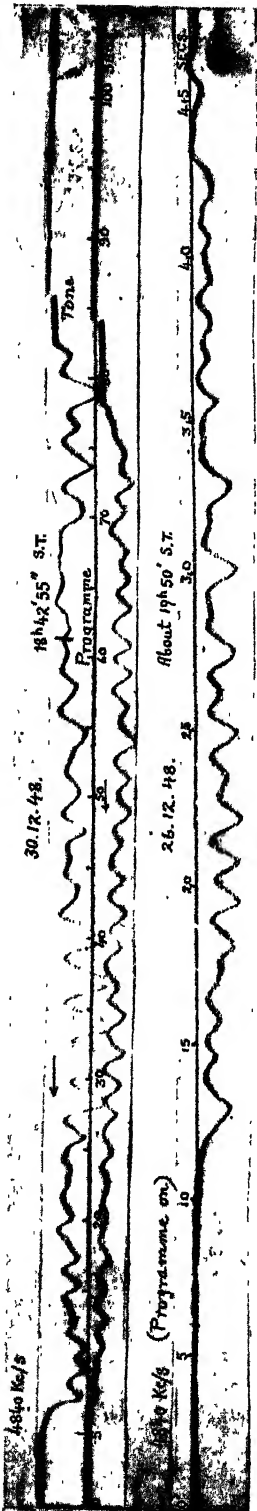


Figure 3. Sinuous patterns with modulated waves.



Figure 4. Slow period fading.

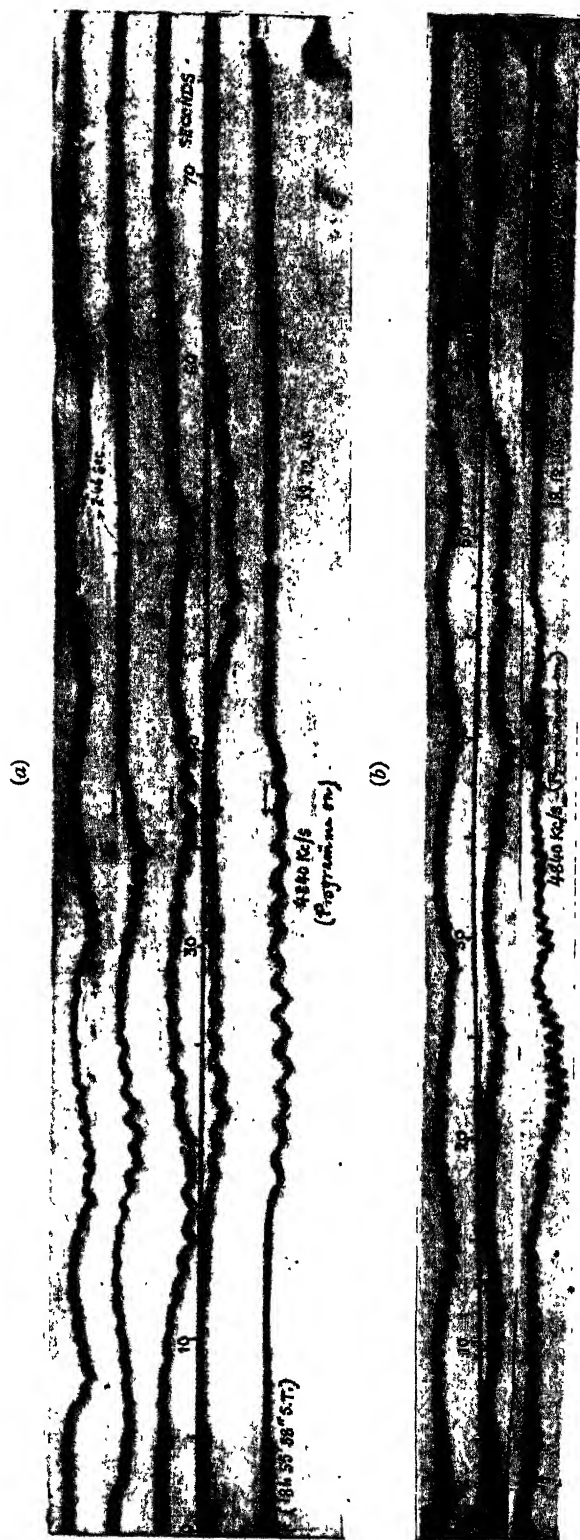


Figure 5. Slow periodic fading with ripples.

where θ_1 and ϕ_1 are the angles of incidence at the E and F layers respectively for simultaneous single reflections.

Considering the signals from Calcutta received at Dacca (distance 240 km.) and taking the E and F2 regions to be at heights of 100 km. and 360 km. respectively, the formula (1) in the case of the singly and doubly reflected waves from the F2 region becomes

$$n \simeq 2v/\lambda \quad \dots\dots (1a)$$

Similarly the formula (2) for the singly reflected waves from both E and F2 regions is reduced to

$$n \simeq 0.62 v/\lambda \quad \dots\dots (2a)$$

If the vertical drift of the ionospheric layer is taken as 4 metres/sec. (about 14 km/hour) during the evening hours, the singly and doubly reflected waves from the F2 region would according to (1a) give periodic fading of nearly 8 cycles/minute whereas according to (2a) the singly reflected waves from both E and F2 regions would produce periodic fading of nearly 2.5 cycles/minute. The observed slow periodic fading (2-10 cycles/minute) can thus be satisfactorily explained according to the Doppler beat interpretation outlined above.

The observed slow periodic fading with superposed ripples was evidently of a mixed type when the ionospheric conditions were favourable for the simultaneous occurrence of the two types of rhythmic fading already described. Thus the quick periodic fading due to the interference between the upper and lower trajectory extraordinary waves in the E layer was found superposed on the slow periodic pattern caused by the Doppler beat effect between the singly and doubly reflected waves from the F2 region or between the singly reflected waves from the E and F2 regions, when the ionospheric conditions were favourable for such effect. It is evident from the formulae (1 a) and (2 a) that the 'beating' of the singly reflected waves from the moving E and F2 regions would cause periodic fading of slightly slower period than the 'beating' of the singly and doubly reflected waves from the same moving F2 region.

The vertical velocity of the ionospheric layer or layers as computed from the observed slow periodic patterns according to the Doppler beat view-point is of the order of 10 km/hour (i.e. about 3 metres/sec.) This agrees with the approximate estimate from the ionospheric data for the observation days and time.

Attention should be directed to the work of Banerjee and his colleagues (1948, 1949) who reported observations of periodic fading with short waves under conditions of relatively high ionospheric ionization. They attributed such periodic fading to the interference between the two waves singly and doubly reflected from the same layer, or between the singly reflected waves from both E and F layers, when there was a continuous change in the path difference of the two interfering waves due to the vertical movement of one or both layers. It is to be noted that this view is essentially the same as the Doppler beat interpretation and that both the views are equivalent representations of the same phenomenon. It can be shown that the same expression is obtained for the periodicity of fading according to the two equivalent conceptions*. For oblique incidence, the frequency of fading can be easily calculated from the Doppler beat formulae (1) and (2).

* See Khastgir and Das (1950).

§ 5. CONCLUSIONS

The main conclusions of the investigation are summarized below :

(i) Of the rhythmic patterns observed during the evening hours at Dacca with the Calcutta signals of 4,840 kc/s. the sinuous pattern of comparatively quick period is considered to be of magneto-ionic origin as being due to the interference of the upper and lower trajectory extraordinary waves in the E layer, the ordinary waves having passed through the E layer.

(ii) The slow periodic fading observed in the evening was caused by the beat effect between the singly and doubly reflected waves from the F2 region or between the singly reflected waves from the E and F2 regions, the two interfering waves having suffered different amounts of Doppler change of frequency due to the vertical movement of the ionospheric layer or layers. The values of the vertical velocity of the ionospheric layers calculated according to the proposed theory appear to accord well with the observed values.

(iii) The slow periodic fading with superposed ripples was observed when the ionospheric conditions were favourable for the simultaneous occurrence of the magneto-ionic type of sinuous fading and the Doppler beat type of slow periodic fading.

(iv) There was evidence of extremely high frequencies (4–12 cycles/second) in a few patterns of periodic fading, the origin of which is unknown.

ACKNOWLEDGMENT

Our sincere thanks are due to Prof. S. K. Mitra, for kindly supplying us with the relevant ionospheric data for Calcutta.

REFERENCES

- APPLETON, E. V., and BEYNON, W. J. G., 1947, *Proc. Phys. Soc.*, **58**, 59.
 BANERJEE, S. S., and MUKERJEE, G. C., 1948, *Phil. Mag.*, **39**, 697.
 BANERJEE, S. S., and SINGH, R. N., 1948, *Indian J. Phys.*, **22**, 413.; 1949, *Science and Culture*, **14**, 293.
 KHASTGIR, S. R., 1949, *Science and Culture*, **15**, No. 3 (Sept.).
 KHASTGIR, S. R., and DAS, P. M., 1950, *Science and Culture*, **15**, 445.
 SUBBA RAO, N. S., and SOMAYAJULU, Y. V., 1949, *Nature, Lond.*, **163**, 442.
 SUBBA RAO, N. S., and SUBRAMANYAM, 1942, *Andhra University Journal*.

Magnetic Hysteresis of Igneous Rocks

By J. McG. BRUCKSHAW AND B. S. RAO

Imperial College, London, S.W.7

Communicated by A. O. Rankine; MS. received 19th May 1950

ABSTRACT. A rapid method has been developed for the investigation of the magnetic properties of igneous rocks. The specimens examined show the normal properties to be expected from a mixture of magnetically different constituents. In particular, high coercivities have been found. The significance of the results in relation to natural permanent magnetism in rocks is discussed.

§ 1. INTRODUCTION

It has been shown that a number of igneous bodies, such as the tholeiite dykes of the North of England (Bruckshaw and Robertson 1949), the Pilansberg dyke (Gelletich 1937) etc., possess permanent magnetism whose direction differs materially from the present direction of the earth's magnetic field. The simplest explanation of this phenomenon is that the ambient magnetic field existing at the time when the bodies cooled through the Curie points of their magnetic constituents, had the direction of the observed permanent magnetism. In the case of the tholeiite dykes, this requires at least a local reversal of the field. Since this does not conform with present theories of the origin of the geomagnetic field, it was necessary to examine the magnetic properties of the rocks to ensure that the adverse magnetism could not arise from some anomalous behaviour of the material. Further, the age of the intrusions varies up to several hundred million years, and to retain adverse magnetism over such a long period a large coercivity is necessary. The present investigation is concerned with the examination of the magnetic hysteresis and a measurement of the coercivity of igneous rocks.

§ 2. PRELIMINARY INVESTIGATIONS

Initially, some observations were made of the variation of the mean volume susceptibility K with different amplitudes of the applied field. For this purpose a method (Bruckshaw and Robertson 1948), previously employed to determine rock susceptibilities in weak fields of about 0.5 oersted, was adapted to measurements up to 50 oersted. In this method an alternating magnetic field is set up by a Helmholtz coil system. In the spatially uniform region at the centre of the system a rock specimen, in the form of a 2 cm. cube, is introduced and, under the action of the applied field, it becomes an alternating magnet which approximates closely to a dipole. When the influence of the applied field is compensated, the E.M.F. set up in an adjacent coil is thus a measure of the induced intensity of magnetization. From this and the known field the mean susceptibility can be determined.

The specimens were samples of tholeiite from the dykes of North England, of dolerite from the Whin Sill and basalts from flows in Nigeria. The Whin Sill, 2×10^8 years old, possesses discordant magnetism (Hallimond and Butler 1949) while the Nigerian basalts also exhibit peculiar magnetism. Several of each type were examined, and typical results of the susceptibility changes up to peak fields of 50 oersted are shown in Figure 1. In the majority of cases the mean susceptibility varied linearly with the peak field, a result consistent with the observations of

Hallimond and Herroun (1943) and the overall change in magnitude was of the order of 10 per cent. The Nigerian basalts, and the samples of tholeiite collected at Swarland, however, gave curves which were not linear and the increase in susceptibility over the whole range of field strengths was large in comparison with the others, in one case as much as 175 per cent. Even in these cases there was no sign of attaining a maximum susceptibility, a feature exhibited by all ferromagnetics.

The absolute values of susceptibility are of little significance, since they merely reflect the content of magnetic material, usually assumed to be magnetite, in the rock. For the tholeiite dykes this content ranged from 3.5 to 4.5 per cent by weight, and for the dolerite about 4 per cent, both values corresponding to about 2 per cent by volume. Slichter (1929) and Nettleton and Elkins (1944) have suggested that

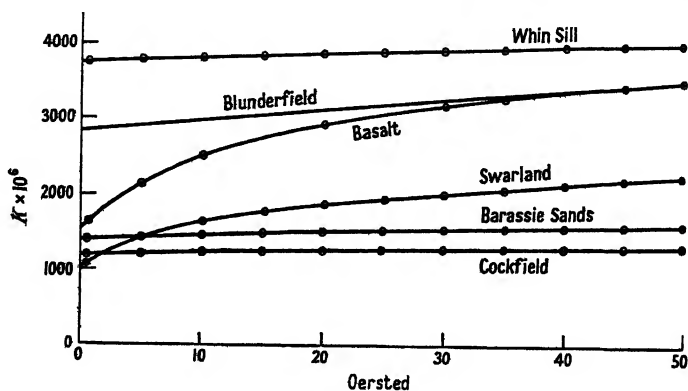


Figure 1.

the susceptibility of a rock containing disseminated magnetite is given closely by the equation $K = 0.3P$, where P is the fraction by volume of magnetite. On the basis of this formula, the present measurements correspond to a magnetite content varying from 0.04 to 0.94 per cent for the tholeiites and from 1.12 to 1.24 per cent for the dolerites. Thus it would appear that the proposed formula is not applicable and it will be shown later that it is impossible for such a simple formula to predict accurate values.

§ 3. THE MEASUREMENT OF HYSTERESIS

Many observers have investigated the magnetic hysteresis of rocks, usually using the well-known magnetometer or ballistic methods. The main difficulty is the small intensities of magnetization, usually about 1 gauss in fields of several hundred oersted, in comparison with those normally encountered in ferromagnetic materials. With the increased sensitivity difficulties arise from instability, particularly in towns. In the present method, the properties are displayed by means of a cathode-ray oscillograph, and the principle of the method is illustrated in Figure 2. An insulating former, of the shape shown, carried five coils of which A_1 , A_2 were used to produce an alternating magnetic field over the specimen S , and the three coils B_1 , B_2 and B_3 to detect the effect of the specimen. These last coils were joined so that the E.M.F.'s generated in B_1 and B_3 opposed that in B_2 . In the absence of the specimen, the number of turns on B_1 and B_3 was adjusted so that the resultant E. M. F. from the three coils was small. To retain a fine control over this adjustment during the experiments, another coil B_3' , of a few turns, was wound over B_3 and joined to a high resistance potentiometer P , from which a small variable

potential difference, of either sign, could be injected into the circuit to obtain the best balance.

When the specimen was introduced, the induction in the volume occupied by it changed from H to $(H + 4\pi I)$ where I is the magnetic moment per unit volume. Since the influence of the time variation in H had been compensated, I only was effective. Because of the different distances involved, the E.M.F. in B_2 due to the changing value of I was greater than the sum of the corresponding E.M.F.'s in B_1 and B_3 , and a resultant E.M.F. proportional to dI/dt was set up across the combination. This E.M.F., which was applied to the Y-sweep of a cathode-ray oscillograph, after suitable amplification, was proportional to $K dH/dt$ where $K (= dI/dH)$ is the incremental volume susceptibility given by the slope of the (I, H) graph. The potential drop across a resistance R , in series with the exciting coils A_1, A_2 , was applied to the X-sweep and, since the field is proportional to the current, yielded a

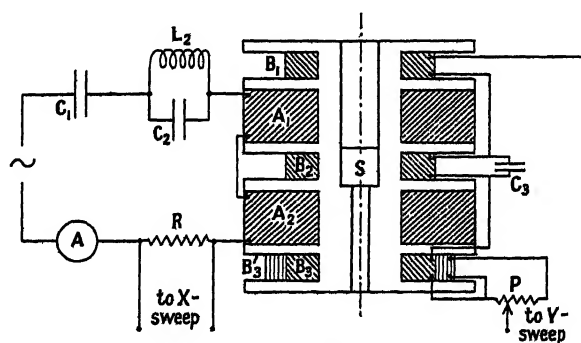


Figure 2.

deflection which was a measure of the instantaneous field. The pattern on the cathode-ray oscillograph was thus a graph showing the variation of $K dH/dt$ with H . With a sinusoidal field $H = H_0 \sin \omega t$, the instantaneous vertical deflection is $KH_0 \cos \omega t$ and, for a constant susceptibility, the pattern would be an ellipse.

§ 4. THE COIL SYSTEM AND THE AUXILIARY CIRCUIT

Details of the coil system are given in Table 1. The inner radius of the coils

Table 1

Coil	Inner radius (cm.)	Outer radius (cm.)	Length (cm.)	No. of turns	Wire (s.w.g.)
A_1A_2	4.2	6.3	2.9	960	20 En. & S.C.
B_3	4.2	—	1.4	1940	40 D.C.C.
B_1B_2	4.2	—	1.4	1850	40 D.C.C.
(approx.)					

was fixed by the size of the rock specimen (a 2 cm. cube) and the strength of the insulating former, while the nearer sides of the exciting coil were controlled by the width of the measuring coil B_2 and the insulating partitions. With these two dimensions fixed, the radial width and length were calculated to give the maximum field per unit current for the length of wire employed. Small departures from the optimum dimensions were not important. Close agreement was obtained between the estimated field and that measured by two different methods. The peak field was 246 oersted per r.m.s. amp. and, with a standard current of 3.05 amp., a

field of 750 oersted was obtained. The uniformity of the field over the space occupied by the specimen was examined and found to be constant to within 2 per cent.

The power supply was a generator, of 532 c/s., driven by a synchronous motor. To obtain the necessary current from the 110 v. supplies, the circuit was tuned by the condenser C_1 . A third harmonic proved troublesome and its influence was substantially reduced by a parallel tuning circuit L_2C_2 adjusted to the harmonic. In balancing the coils B_1, B_2, B_3 , their self-capacities gave small phase displacements between the potential differences across their terminals, and it was necessary to place a small variable condenser C_3 across one coil to obtain a suitable balance. The

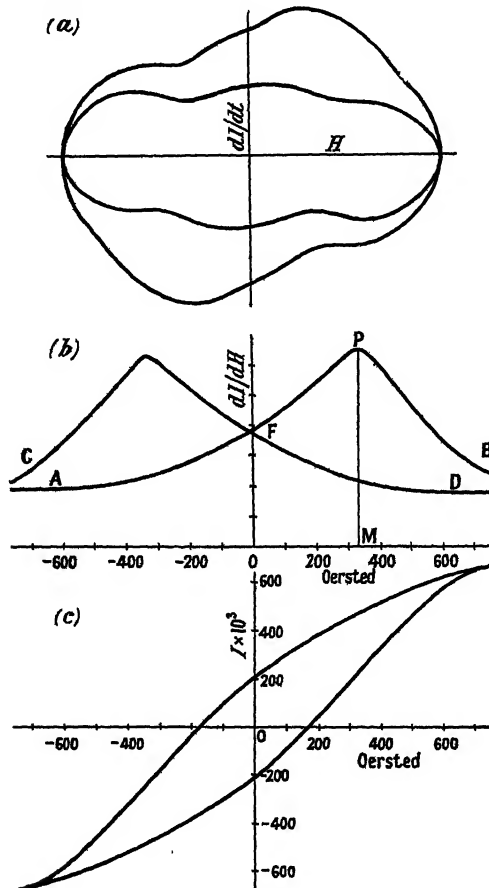


Figure 3.

adjustments were correct only for one frequency and a residual pattern was left due to imperfectly balanced harmonics. As this pattern intersected itself it was found convenient to open it out by adjusting the potentiometer P. Frequently, to accommodate large deflections due to the specimen, the initial trace was adjusted so that the specimen caused the trace to collapse and open out in the opposite direction.

A photographic record was made for each specimen, the record consisting of the trace without the specimen, the trace with the specimen in position and the two axes, the last being obtained by disconnecting first the X-sweep and then the Y-sweep. A reproduction of a typical record is shown in Figure 3 (a).

§ 5. ANALYSIS OF THE RECORDS

Subtracting the deflections obtained with and without the specimen gave values proportional to $KH_0 \cos \omega t$ for various values of the applied field. The latter value was known in oersted since the maximum X-deflection corresponded to the peak field of 750 oersted. As the combination of amplifier and cathode-ray oscillograph did not give deflections strictly proportional to the input, corrections were necessary to take this into account. For any given value of H , the value $\omega t = \sin^{-1}(H/H_0)$ could be determined and hence the Y-deflections converted into values proportional to K . These values were plotted, yielding curves of the type shown in Figure 3 (b), in which the curve AB corresponds to the field increasing (upper part of the original record) and DC to the field decreasing (lower part of record). As would be expected, the latter is the mirror image of the former in the Y-axis. It should be noted that the susceptibilities at the maximum and minimum fields are extrapolated, since both $\cos \omega t$ and the deflection are zero here and the ratio indeterminate.

Numerical integration of the $(dI/dH, H)$ curves yielded the shape of the (I, H) loop with the intensity in arbitrary units. Further, since the integration must commence at some arbitrarily selected zero, the position of the curve in relation to the axis of intensity of magnetization was unknown. In the cyclic condition, it was assumed that the intensities for fields $+H_0$ and $-H_0$ were equal and opposite, and the line $I=0$ could then be fixed. The intersection of the curve with this line permitted the coercivity to be determined. Integration round a complete cycle was used to obtain some idea of the accuracy of the results. In the example shown in Figure 3 (b), the integration of the curve AB gave an overall change of 45.1 arbitrary units of intensity and the branch CD a change of 44.1 units. The former half cycle yielded a coercivity of 183 oersted and the latter 180 oersted. Discrepancies of a similar magnitude were found in other cases, and the mean coercivity quoted here is probably accurate to 2 or 3 oersted. Some idea of the magnitudes of the susceptibility and of the intensities of magnetization can be obtained by assigning to OF of Figure 3 (b) the susceptibility observed in the preliminary experiments using a small field. The intensities shown in Figure 3 (c) were derived on this basis. This substitution, obviously, is not strictly correct, but the error involved cannot be serious. Indeed, for the present purpose the absolute values, reflecting the concentrations of magnetic materials, are of little significance. A direct calibration of the system is possible, for by using the equivalent of a current-carrying loop and a magnetic shell, a coil wound with n turns per unit length corresponds to a volume intensity of magnetization of ni , where i is the current in E.M.U.

§ 6. EXPERIMENTAL RESULTS

The rock samples examined in the preliminary tests were also examined in the stronger fields. From the measurements two values characteristic of the specimen, could be determined, the value H_m of the field corresponding to the maximum susceptibility and H_c , the coercivity. The ratio of the maximum susceptibility to the value at $H=0$, i.e. PM/OF in Figure 3 (b), should also be characteristic but as this varied little, from about 1.5 to 1.7, it was not used. The results for most rocks are shown in Table 2. Examination of Figure 3 (b), in which the susceptibility does not fall to zero at high fields, and which also exhibits a discontinuity on reversal, shows that the specimen did not attain saturation*. This was true in all

* On the basis of disseminated spheres of magnetite an estimated field of about 2,200 oersted would be required for saturation.

cases, and the coercivities given here are underestimates of the true coercivities. There is a corresponding uncertainty in H_m . Nevertheless, as each rock was examined under standard conditions a comparison of the results is valuable.

The results of Table 2 refer to those specimens which gave simple curves. The tholeiite specimens from the Swarland exposure and the Nigerian basalts, however,

Table 2

Specimen	H_m (oersted)	H_c (oersted)	$K_0 \times 10^6$
THOLEIITES			
<i>Aklington Dyke</i>			
Abington exposure	338	165	1920
	375	180	1960
	263	120	2640
Lugton exposure	255	113	1920
Cockfield exposure	319	195	980
<i>Cleveland Dyke</i>			
Blunderfield exposure	223	113	1290
	338	182	1050
	263	165	2000
Hazel Cottage exposure	338	180	2000
	450	270	1400
DOLERITE			
Whin Sill at Cross Fell	300	188	3520
	300	210	3400

gave anomalous results, the former exhibiting two maxima on the susceptibility curve and the latter three maxima. The original cathode-ray oscillograph pattern for a sample from the Swarland exposure is reproduced in Figure 4. The results for the anomalous specimens are summarized in Table 3.

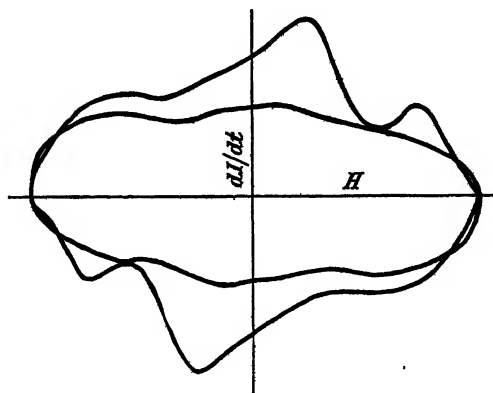


Figure 4.

§ 7. DISCUSSION OF RESULTS

The most obvious interpretation of the multiple maxima on the susceptibility curve is the existence within the rock of two or three magnetic materials with distinct properties. This is consistent with observations, made at the Imperial College and elsewhere, on the variation of susceptibility with temperature in which a specimen may show several Curie temperatures distinct from the generally

accepted value of 580°C . for magnetite. For specimens with more than one component, the observed values of H_m and H_c will not have any precise significance. Since the peak due to one component is superimposed on a rising or falling background due to the other materials, a displacement of the peak will occur. The coercivity, on the other hand, will be intermediate between the greatest and least coercivities of the constituents. During the demagnetization of the specimen, when the field attains the smallest coercivity, some components will still retain their magnetism while, when the greatest coercivity is reached by the field, other components will possess magnetism in the direction of the field. The condition that the average intensity of magnetization is zero will be attained between these two extremes, at a field strength dependent on the relative concentrations and the intrinsic properties of the individual constituents.

The results in Table 2 in which the specimens appear to show only one magnetic component, demonstrate that there is little constancy in either H_m or H_c , the former ranging from 223 to 450 oersted and the latter from 113 to 270 oersted. Even in samples from the same exposure, considerable variations are apparent, as illustrated by the three collected at Abington.

Assuming that the properties of the magnetic constituents are similar to those of magnetite, the order of magnitude of H_m is to be expected. Specimens of natural magnetite give maximum susceptibilities at fields lying between 20 and 50 oersted,

Table 3

Specimen	H_m (oersted)	H_c (oersted)	$K_0 \times 10^6$
<i>Tholeiites</i>			
Swarland exposure	173, 563	98	720
	280, 660	189	500
<i>Basalts</i>	137, 516, 720		600
Nigeria	208, 572, 675		900
	150, 525, 675		640

when they possess intensities of magnetization of the order of 50 gauss. Assuming in the rock spherical particles, with a demagnetization factor of $4\pi/3$, an external field of 260 oersted is necessary to produce an internal field of 50 oersted. Fluctuations in the mean demagnetization factor for the particles would account for some of the variation in the value of H_m . This, however, would appear to account for only a small part of the observed range for, when $I=0$, the internal and external fields are equal. Thus, if all of the disseminated particles had the same magnetic properties, the rocks should possess the same coercivity. There is thus clear evidence that a wide variety of magnetic properties is involved, even in those specimens which, by virtue of a single peak, appear to contain one constituent only. Such a variety may be attributed to different impurities, particle size and heat treatment. The last factor is necessarily connected with the second. Basalt flows in general cool rapidly and show a fine grained, and even vitreous, structure. There is also the possibility of reheating by subsequent flows over them. Sills and dykes, on the other hand, will cool more slowly, but the reduction of temperature will be faster in the parts in contact with the country rock than in those near the centre. Such a differentiation of cooling may well play a part in accounting for the variations in samples from the same exposure.

Although it seems probable that there is a continuous range of properties arising from different factors, it must not be assumed that each specimen in Table 2 contains only one type of material because it exhibits one maximum value of

susceptibility curve. This curve will only resolve the influence of the individual constituents if the peaks are separated by an interval greater than a critical amount. At smaller intervals the two will merge into a single broad peak. The critical interval is obviously related to the relative magnitudes of the peaks and their shape.

It is apparent that in these conditions the simple formula, suggested by Slichter for the susceptibility in weak fields in terms of magnetic content, is untenable. If the range of properties is due to a finite number of constituents, it might be possible to replace it by a sum of similar terms, each relating to a single constituent. Even this seems improbable and the formula must be regarded as an over-simplification.

The coercivity values which have been determined are, for reasons given, underestimates of the greatest coercivity involved. Even so, they are high in comparison with the values for iron, cobalt, nickel etc. for which the values range from 1 to 12 oersted. The present values are comparable with those of modern magnet steels. They are similar to those obtained by other investigators, such as Puzicha (1930) who obtained values from 5 to 226 oersted. If the permanent magnetism of the tholeiite dykes was, as suggested, acquired shortly after their injection, it has been retained for 20 to 30 million years against a demagnetizing external field for a large part of this period. The Whin Sill, on the same assumptions, has retained its magnetism for several hundred million years, as also must the Pilansberg Dykes of S. Africa. It is difficult to convert the coercivities as measured into loss of magnetic moment with time. Many factors may assist in demagnetization such as earth tremors, stresses due to earth movements, temperature changes etc. Nevertheless, the high coercivities make reasonable the assumption that, if the earth's field has not changed appreciably in magnitude, some of the original intensity of magnetization would be retained.

Finally, it is obvious that the magnetic behaviour of these rocks is normal and it does not display any anomaly which might account for the observed adverse polarization of many igneous bodies.

ACKNOWLEDGMENT

The authors wish to record their thanks to the Anglo-Iranian Oil Company whose financial assistance made this investigation possible.

REFERENCES

- BRUCKSHAW, J. M., and ROBERTSON, E. I., 1948, *J. Sci. Instrum.*, 25, 444; 1949, *Mon. Not. Roy. Astr. Soc., Geophys. Suppl.*, 5, 308.
 GELLETICH, H., 1937, *Beitr. Geophys.*, 6, 337.
 HALLIMOND, A. F., and BUTLER, A. J., 1949, *Mining Mag.*, 80, 265.
 HALLIMOND, A. F., and HERROUN, E. F., 1943, *Proc. Phys. Soc.*, 55, 214.
 NETTLETON, L. L., and ELKINS, T. A., 1944, *Geophysics*, 9, 60.
 PUZICHA, K., 1930, *Z. prac. Geol.*, 38, 161, 184.
 SLICHTER, L. B., 1929, *Geophysical Prospecting, Amer. Inst. Mech. Engng.*, New York, 238.

The Michelson Interferometer at Millimetre Wavelengths

By W. CULSHAW

Telecommunications Research Establishment, Ministry of Supply, Great Malvern, Worcs.

Communicated by R. A. Smith; MS. received 27th April 1950

ABSTRACT. The design and operation of an interferometer of the Michelson type at a wavelength of 12.5 mm. is discussed.

The required frequency stabilization of the 12.5 mm. source is achieved by using a high Q cavity as an R.F. discriminator.

Wavelength measurements have been made with various spacings of the interferometer, the measured wavelength increasing as the spacing is reduced. Results indicate that the interferometer gives more accurate measurements of the wavelength when operated well in the Fraunhofer region of diffraction.

The operating frequency has been measured using a calibrated frequency meter, and hence the velocity of electromagnetic waves deduced. The value obtained for this velocity agrees, within the accuracy attempted with the present instrument, viz. one part in 10^4 , with the generally accepted value for this velocity. The possibilities of increasing the accuracy of its determination with the interferometer are discussed.

Measurements of the dielectric constants of low-loss materials by means of the interferometer have been made. Results obtained using ordinary commercial sheets of materials agree within a few per cent with values obtained otherwise.

The use of the interferometer as a substandard of length is also discussed.

§ 1. INTRODUCTION

IN order to ensure that propagation occurs only in the dominant or fundamental mode, the lateral dimensions of conventional transmission systems, such as waveguides, are made smaller than the operating wavelength. Consequently for operation at millimetre wavelengths the dimensions of waveguides are small and the attenuation is large. For a similar reason the dimensions of components such as cavity wavemeters are small in the millimetre wavelength region, and the resulting increased attenuation adversely affects their discrimination. In addition there is a decrease in the tolerance permissible in the manufacture of such components. It thus becomes apparent that for these small wavelengths alternative forms of transmission systems and measuring instruments are desirable, whose lateral dimensions are not limited by the wavelength. In this connection it is natural to consider the application of optical techniques to millimetre wavelengths.

The principle of any optical instrument can be applied to millimetre wavelengths. It cannot be assumed, however, that the resulting millimetre wave instrument will perform its function as efficiently as its optical counterpart; the effect of the difference in wavelength must be considered, both on the technique of the particular measurement and on the physical properties of materials used in the construction of the instruments. The most important limitation in this connection is the diffraction which occurs at millimetre wavelengths with radiating apertures of practical dimensions. With the generators available at present the aperture dimensions are restricted to about $\cdot 15$ wavelengths. The beams obtained are consequently divergent, the constant phase fronts being by no means plane, and dependent on the distance from the radiating aperture.

The problems of obtaining an efficient transmission system and precise measuring techniques at millimetre wavelengths are simpler, as far as applications are concerned,

techniques is concerned, the smaller the wavelength. As in many optical instruments, for instance the optical telescope, there is also the possibility of screening the system, but again it is necessary to consider the effect of the difference in wavelength. In the millimetre wavelength region with practical tube sizes the wavelength in the tube depends on its dimensions and the free space wavelength is not obtained. In addition there is the possibility of exciting other modes of propagation in such tubes with a consequent loss of energy and discrimination.

In order to study some of the problems in this field the principles used in the Michelson interferometer have been adapted for operation at a wavelength of 12.5 mm. and an instrument has been constructed. This paper describes the various factors involved in the design and operation of the instrument together with the results of measurements made with it.

§ 2. DESIGN CONSIDERATIONS

2.1. *General Principles*

The optical form of the Michelson interferometer is well known (Wood 1934). Light from an extended monochromatic source is divided by a 'half-silvered' glass plate and caused to travel different paths at right angles to each other. The beams are reflected by mirrors and combine at the observer. Due to the multiple reflections in the interferometer, the mirrors can be regarded as being on the same straight line and separated by a distance t , where t is the difference of the distances of the mirrors from the 'half-silvered' glass plate.

If the mirrors are parallel then we have a maximum intensity when

$$2\mu t \cos \theta = n\lambda, \quad \dots\dots(1)$$

where λ is the free space wavelength, μ is the refractive index, and $n=0, \pm 1, \pm 2$ etc. We have a minimum intensity when

$$2\mu t \cos \theta = (2n+1)\lambda/2; \quad \dots\dots(2)$$

$n=0, \pm 1, \pm 2$ etc.

On focusing the eye at infinity a system of circular fringes is seen, since θ the angle of incidence varies when an extended source is used. In general, therefore, complete darkness is not obtained, the fringe system, or energy distribution across the aperture merely changing with distance t . The intensity variation of the fringes is of the form of $\cos^2 \phi$, the maxima and minima diffusing into each other.

For operation in the millimetre wave region transmitting and receiving apertures are used in place of the extended light source and the 'eye'. Since energy is received over the area of the receiving aperture, this necessitates the use of essentially parallel beams in the interferometer in order to obtain sharp maxima and minima (i.e. it is desirable to work with a single fringe). It is also essential to consider diffraction effects in relation to the size of apertures and the spacing of the component parts of the interferometer.

A difficulty encountered in the theory of diffraction of electromagnetic waves is the determination of the field in the radiating aperture consistent with the boundary conditions. The most familiar theory is that of Huyghens-Kirchhoff (Stratton 1941) which assumes that the field in the aperture is that which would exist if the conducting part of the screen were not present.

This theory gives good approximations for apertures whose dimensions are large compared with the wavelength of the radiation, as is usually the case in optics where the Huyghens-Kirchhoff formula forms the whole basis of diffraction

theory. The more familiar Fraunhofer and Fresnel types of diffraction follow from this general formulation.

At millimetre wavelengths the aperture is not always large compared with the wavelength, and some care is necessary in the application of the above formula. However, it has a wide application in the field of aerial theory with aperture dimensions of the order of 10 wavelengths, and there is reason to suppose that it will give reasonable practical results for both relatively near and distant diffraction patterns when the dimensions of the aperture are of this magnitude.

A consideration of the diffraction patterns at various distances from a long slit some 15 wavelengths wide at millimetre wavelengths leads to the following conclusions, the slit being supposed uniformly illuminated by a plane wave front.

(i) If λ is the free space wavelength, a the width of the slit, and z the distance along the normal to the slit, then when $\lambda z > a^2$ Fraunhofer diffraction methods can be employed. These give the diffracted field at relatively large distances from the radiating aperture. For this region, near the z axis through the centre of the slit, the constant phase fronts are essentially spherical, the curvature decreasing as z increases, and the amplitude of the diffracted field is inversely proportional to z .

(ii) For $\lambda z < a^2$, but not below a certain value which depends on the obliquity factor and the inverse variation with distance of the amplitude of the Huyghens' sources, Fresnel diffraction methods are used. In this region, on planes perpendicular to the z axis, the amplitude of the diffracted field fluctuates through maxima and minima, decreasing rapidly as the geometrical shadow is approached but extending beyond it. The shapes of the amplitude and phase diagrams of the diffracted field vary rapidly with distance z , the constant phase fronts being by no means plane.

(iii) The distance z_0 given by $\lambda z_0 = a^2$ may be called the transition distance between the regions of Fraunhofer and Fresnel types of diffraction for the particular case here considered. The criterion should be applicable to rectangular and circular apertures provided the appropriate width is substituted, the apertures again being uniformly illuminated by a plane wave front. The criterion gives the minimum distance at which the usual polar diagram of such radiating apertures may be measured, and for a given wavelength is directly proportional to the gain of the aperture.

Ideally the interferometer requires plane wave trains for its operation in the millimetre wave region, but since this is not practicable it is more likely to function most efficiently in a region of uniform diffraction such as in the Fraunhofer region, rather than in the region of Fresnel diffraction where violent changes occur in the amplitude and phase of the diffracted field. In order to preserve a balance between the two beams as mirror 1 is moved over its traverse, and also to obtain adequate energy transfer through the interferometer, the divergence of the radiated energy must be small. The practical size of the apertures used is thus a compromise between the beam divergence and the distance from the radiating aperture to the Fraunhofer region of diffraction, since the larger we make this aperture the greater this distance becomes. The transmitting and receiving apertures in the present interferometer are 6 in. \times 6 in., the linear dimensions thus being about 12 wavelengths of 12.5 mm. An aperture of these dimensions supposed uniformly illuminated has an angle of 10° between the first zeros either side of the

of the Fraunhofer diffraction pattern, and the distance z_0 from the aperture to the Fraunhofer region as given by the above considerations is about 180 cm. This means that for operation in the Fraunhofer region the two path lengths in the interferometer between the transmitting and receiving apertures should both be greater than 2 metres.

2.2. *Transmitting and Receiving Apertures*

The transmitter consists of a reflection klystron valve generating radiation of wavelength about 12.5 mm. Power from this is fed into a rectangular waveguide of sectional dimensions 0.170 in. \times 0.420 in., and the waveguide is then gradually enlarged in the form of a pyramid to form a horn with aperture 6 in. \times 6 in. In order to approach the condition of uniform illumination, and thus to reduce the divergence of the transmitted beam, the horn is terminated by a metal lens of the appropriate focal length. This lens transforms a spherical wave diverging from the point of intersection of the sides of the pyramid into a plane phase front at the radiating aperture.

The receiving aperture consists of a similar horn which is tapered to the size of the waveguide, the received signal being fed into a millimetre-wave crystal detector. The receiving horn is also phase corrected by a metal lens in order to increase its effective receiving area. The increase in size from the waveguide of sectional dimensions 0.170 in. \times 0.420 in. to the 6 in. \times 6 in. aperture takes place over a length of 50 cm., and the metal lenses employed have focal lengths of 52.5 cm. The design of these 'metal lenses' is now well known (Kock 1946). The fact that the phase velocity of electromagnetic waves in a waveguide is greater than in free space is utilized to simulate a medium of the appropriate refractive index. The profile of the lens is then calculated using geometrical optics. The metal lenses used have a refractive index of 0.6 and hence a plate spacing of 7.81 mm., and they have proved very effective in obtaining an efficient transfer of energy through the interferometer. Figure 1 shows a schematic diagram of the millimetre-wave interferometer and Figure 2 is a photograph of the equipment used.

2.3. *Beam Divider*

In the optical form of the Michelson interferometer a 'half-silvered' glass plate is used to produce the two beams required; ideally this 'beam divider' sends 50% of the incident energy along each path. In view of the fact that silver is almost a perfect reflector at millimetre wavelengths, this technique did not seem practicable, and other forms of 'beam divider' have been used. Experiments were made using various types such as wire and slot gratings, designed to reflect 50% of the energy incident at 45°. These gratings have operated quite well in the interferometer, but were superseded by a more efficient and adjustable reflector which consisted of two quarter-wave plates of dielectric with an adjustable air gap between them as shown in Figure 1. The plates are quarter-wave thick for 45° incidence and their separation is adjusted to give a power reflection coefficient of 50%. This technique has proved satisfactory using polystyrene plates of dielectric constant about 2.5. The adjustable dielectric beam divider has been useful in another way. If the wave fronts in the interferometer were plane, the balance at a minimum would not depend on the reflecting properties of the 'beam divider', since there is a transmission and a reflection for each path. Thus the balance of the two beams would be unaffected

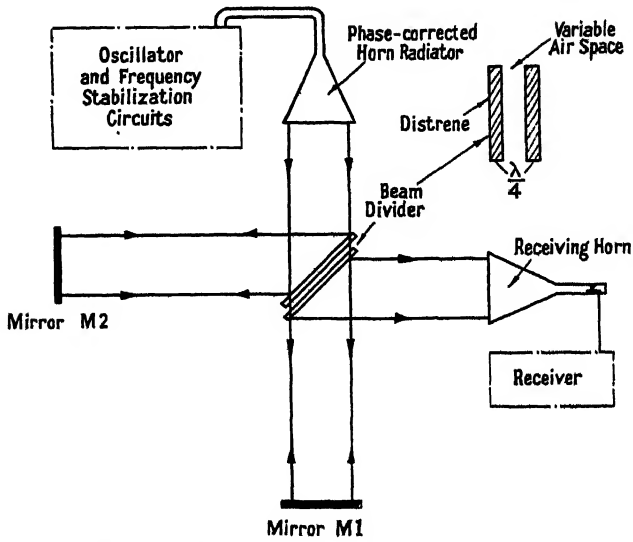


Figure 1. Millimetre-wave Michelson interferometer.

by the value of the reflection coefficient, although the sensitivity and energy transfer from transmitter to receiver would be greatest for a power reflection coefficient of 50%. However, in the present interferometer the beams are divergent, and any asymmetry in the two paths will destroy the balance, but since the reflectivity of the dielectric 'beam divider' can be adjusted to compensate



Figure 2. Millimetre-wave Michelson interferometer.

for this, a balance of the beams can be obtained when their path lengths are different. This facility has been useful in obtaining adequate sensitivity in the interferometer. In view of the fact that measurements were contemplated at various spacings of the interferometer it did not seem necessary to use a beam divider much larger than the radiating aperture. The dimensions of the divider used are 9 in. wide and 8 in. high.

2.4. Mirrors M1 and M2

The mirrors are made slightly larger than the apertures in order to reduce possible 'edge effects'. They have dimensions 8 in. \times 8 in. with a surface tolerance of 0.0005 in., optical tolerances, of course, being unnecessary at these

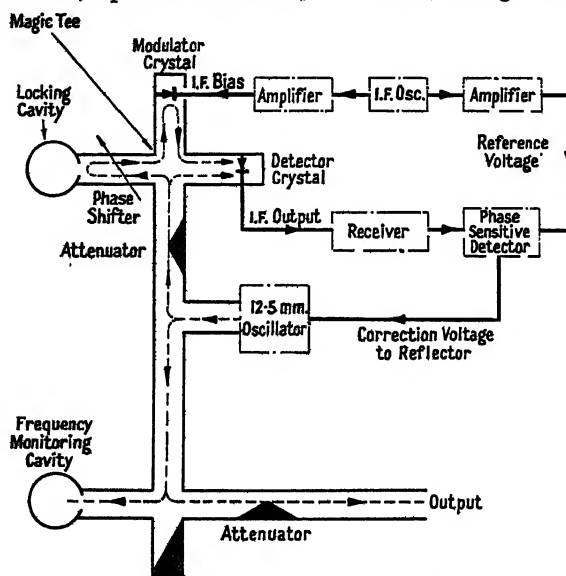


Figure 3. Block diagram of frequency stabilization system used on interferometer.

wavelengths. The mirrors are machined from brass which is then silver-plated. One of the mirrors, M1, is mounted on a V-slide and is movable by means of a lead screw of 2 mm. pitch over a distance of 50 cm., with a reading accuracy better than 0.0005 cm.

2.5. Frequency Stabilization

From observations made using a cavity wavemeter to measure the frequency stability of the 12.5 mm. source, it became clear that if precise measurements of wavelength were to be made by means of the interferometer, then some frequency stabilization of the source was necessary. A high degree of frequency stability has been achieved using the Pound Frequency Stabilizer (Pound 1946), in which the 12.5 mm. source is effectively locked to a cavity of high Q . The system employed in the interferometer is shown in Figure 3, an additional cavity with a Q value of some 18,000 being provided for frequency monitoring. The frequency of the source is locked to the frequency controlling cavity, which is then tuned to the frequency monitoring cavity so that the operating frequency is kept constant.

The frequency stability of the present system is about one part in 10^6 for c.w. operation, and could be improved by suitable thermostatic control.

§ 3. OPERATION OF THE INTERFEROMETER

3.1. *General Considerations*

Apart from the difficulties due to diffraction, the Michelson interferometer at millimetre wavelengths operates with essentially parallel beams, so that the circular fringe system of the optical model is not obtained. Instead, interference occurs at the receiving horn between two sinusoidal wave trains which have traversed different paths, producing essentially a single fringe. When the path difference is such that a minimum is obtained, there is an effective short circuit in the receiver arm, and ideally all the radiated energy is reflected back into the transmitting horn.

When mirror M1 is moved, the received signal fluctuates through maxima and minima in much the same way as in a standing-wave pattern on a short-circuited transmission line. Since the sharpness of the minimum depends on the equality of two beams which have traversed different paths, this sharpness will vary as M1 moves along its traverse, due to diffraction and scattering effects, so that the high maximum to minimum ratio of a short-circuit transmission line cannot be obtained for all positions of M1.

As might be expected, the setting-up of the millimetre-wave interferometer is not as critical as the optical model. The slide of mirror M1 is first adjusted by means of the screws provided so that M1 moves along the axis of the transmitting horn, the mirror M1 already being fixed in the plane perpendicular to its slide. This is best done visually, or using the beam divider to reflect the energy returned from mirror M1 into the receiving horn mounted at right angles to this axis, the slide is adjusted for maximum received signal when mirror M1 is at either end of its traverse. Mirror M2 is now placed in position and adjusted for a maximum received signal, both mirrors being at the same distance from the beam divider.

Adjustment of the beam-divider spacing for maximum received signal is also made at this stage. Finally, slight adjustments of the mirrors and beam divider are made to give adequate sharpness to the minima.

3.2. *Measurement of Wavelength and Frequency*

The apparent wavelength of the radiation as measured with the millimetre wave interferometer is obtained by noting the readings of minima which occur as M1 is moved over its traverse. The distance between adjacent minima should be half a wavelength, so that by observing the distance over which M1 moves for a given number of minima the apparent wavelength at the particular spacing of the interferometer can be determined. Once the apparent wavelength has been measured the interferometer can be used to measure arbitrary distances.

Additionally the frequency of the radiation can be measured by means of a cavity wavemeter calibrated by reference to a crystal-controlled oscillator. In the present experiment the accuracy of this calibration was limited by the setting accuracy of the wavemeter to be ± 1 Mc/s. at 24,000 Mc/s. This accuracy could be improved, but it was considered adequate for the present experiments. From a measurement of the frequency made in this way, and the wavelength as determined by the interferometer, the velocity of electromagnetic waves in air can be computed.

In the actual determination of wavelength 80 minima are counted as M1 moves over its traverse of 50 cm. The positions of the minima at the ends of this traverse are determined either by setting on the actual minimum or by setting slightly off the minimum on either side of the response where a finite

exists. In the latter the setting accuracy is better, but care must be taken that the response is symmetrical or erroneous results may be obtained.

Using transmitted powers of a few milliwatts, and with the component parts of the interferometer spaced at distances not exceeding 2 metres from the beam divider, the measurement of the rectified current through a 12.5 mm. crystal detector in the waveguide feed from the receiving horn has proved adequate, and currents from 200–400 microamps have been obtained. Maximum to minimum ratios of this current up to 10,000:1 have been noted in some instances. Due to diffraction effects there is, as mentioned earlier, some variation in the maximum to minimum ratio of this current as the mirror M1 is moved. Nevertheless, using a spot galvanometer of 20 ohms resistance and a full-scale deflection of 8 microamps, setting accuracies on the minimum within 10^{-3} cm. can be achieved in most instances when using the 6 in. \times 6 in. transmitting and receiving apertures; by setting on the sides of the response the setting accuracy approaches 10^{-4} cm., the limit being set by slight power output fluctuations of the oscillator value.

For wider spacings of the interferometer the sensitivity using the crystal detector and galvanometer technique is inadequate, and a 12.5 mm. super-heterodyne receiver is used. The local oscillator of this receiver is locked to a cavity by means of another Pound Frequency Stabilizer, in order to keep the receiver in tune. In spite of the large gain available in such a receiver it has not been found possible to improve on the setting accuracy obtained with the crystal and galvanometer technique. This is due to the effects of diffraction and stray radiation in the present interferometer.

3.3. *Measurement of Dielectric Constants*

Using the optical analogy it would appear that the dielectric constant of a plane sheet of low-loss material is determined simply by inserting a sheet of the material normally into one of the beams of the interferometer and measuring the resultant shift of a minimum. The ambiguity as regards the number of minima in the actual shift would have to be resolved by using sheets of different thickness, unless the magnitude of the dielectric constant is known approximately. The refractive index, and hence the dielectric constant of the material, is given by

$$t(\mu - 1) = x, \quad \dots\dots(3)$$

t being the thickness of the sheet, μ the refractive index, and x the shift of a minimum. This method, even neglecting diffraction effects, can only give an approximate result, since it neglects the reflections which occur at the surfaces and inside the specimen; it will thus give better approximations for low dielectric constants.

Whilst approximate results using this method have been obtained for Perspex and polystyrene sheets, the values of dielectric constant obtained have been found to depend on both the size and the position of the specimen in the beam, this dependence being no doubt connected with the diverging nature of the beams in the interferometer as well as the reflections occurring at the specimen.

To eliminate these uncertainties the sheets were made the same size as the mirrors and were mounted directly on to mirror M1. In this position the method becomes analogous with the short-circuited line technique (von Hippel and Roberts 1946) of determining dielectric constants.

The simple formula of equation (3) gives the first approximation, but even with dielectric constants of the order of two, more accurate methods of impedance transformation must be applied. Briefly, measurements of the shift of a minimum on placing the specimen on mirror M1, and the thickness, enable a determination to be made of the distance of the first minimum in air from the dielectric boundary. Then by applying impedance transformations back from this minimum to the dielectric-air surface, and also from the mirror through the dielectric to this interface and equating the two impedances, we obtain a transcendental equation involving the dielectric constant as an unknown. In practice, a simpler method for dielectric materials of low loss is to use the circle diagram of impedances and to perform this matching of impedances by trial and error, using the approximate value for the dielectric constant, equation (3), as a guide.

§ 4. RESULTS

4.1. *Spacing of the Interferometer*

Measurements have been made at five positions of the interferometer. The respective distances in centimetres of the component parts from the beam divider are shown in Table 1.

Table 1

Position	1	2	3	4	5
Transmitting horn	150	100	50	25	310
Receiving horn	150	100	50	25	150
Mirror M1	175	125	75	50	600
Mirror M2	175	125	75	50	40

The distance of mirror M1 from the beam divider was measured when M1 was at the centre of its traverse, i.e. when set at 25 cm. on the 50 cm. scale.

4.2. *Wavelength Measurements*

Measurements of the apparent wavelength were made at each of the positions given in Table 1, using the 6 in. \times 6 in. apertures of the receiving and transmitting horns. These measurements were made both by setting on the actual minima and by setting on the sides of the response. The results are shown in Table 2. Deviations from the mean of the measurements made in each position were about one part in 10^4 . During all determinations of the wavelength the operating frequency was constant at $24,070 \pm 1$ Mc/s., as indicated by the calibrated wavemeter. The lead screw used to move mirror M1 has been checked at 20° C. and found to have an average error of one part in 10^4 over a length of 40 cm. This error has been allowed for in the results quoted in Table 2.

Table 2. Apparent Wavelength

Position	1	2	3	4	5
Wavelength { using minimum	1.2452	1.2455	1.2456	1.2459	1.2451
(cm.) { using response	1.2452	1.2455	1.2456	1.2459	1.2451

Values of the apparent velocity of electromagnetic waves in air at 20° C. have been deduced from the known frequency and measured wavelength of the radiation, and the results are shown in Table 3.

Table 3. Measured Velocity of Electromagnetic Waves.

Position	1	2	3	4	5
Velocity (10^{10} cm/sec.)	2.9972	2.9980	2.9982	2.9989	2.9970

Using the present standard value for the velocity of light *in vacuo*, viz. 2.99776×10^{10} cm/sec., and assuming a value 1.0006 for the dielectric constant of air under the conditions of the experiment, we obtain for the velocity of electromagnetic waves in air a value of 2.9969×10^{10} cm/sec. to the approximation required, viz. 1 in 10^4 . To this approximation the effects of variations in humidity and pressure on the measurements have been neglected. These would need consideration in a more accurate determination.

From Table 2 it is seen that the value obtained for the wavelength depends on the spacing of the component parts of the interferometer. The measured wavelength increases as the spacing is reduced, this being due no doubt to the spherical wave fronts existing in the interferometer. From Table 3 it is seen that

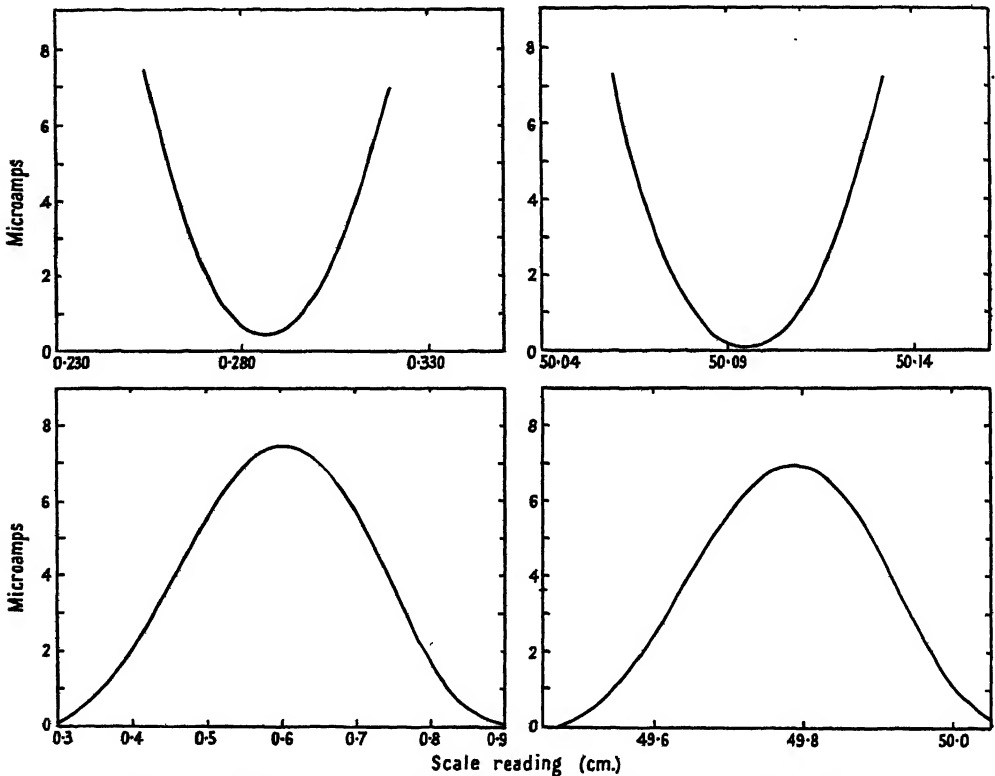


Figure 4. Maximum and minimum responses at position 1; 6 in. \times 6 in. apertures.

there is agreement, within the limits of experimental error, between the deduced velocity of electromagnetic waves and the generally accepted value, this agreement occurring at the wide spacing of the interferometer in position 5. The operating wavelength as measured in position 5 is 1.2451 ± 0.0001 cm.

Measurements of the responses at a maximum and minimum were made with the interferometer in position 1. The results are shown in Figure 4. The symmetry of the minimum response approaches 1 in 10^4 , which is about the accuracy of the wavelength determination in position 1.

Figure 5 shows a plot of a number of maxima and minima occurring on the interferometer around a scale reading of 25 cm. for mirror M1, the mirrors then being at equal distances from the beam divider,

Maximum and minimum responses with the interferometer in position 4 are shown in Figure 6. The symmetry of the minimum response no longer approaches 1 in 10^4 , and the accuracy of the wavelength measurement at this position is no longer within this limit.

The effect of reducing the 6 in. \times 6 in. radiating and receiving apertures was studied in position 4. The apertures were reduced by mounting $\frac{1}{8}$ -in. thick brass plates over the ends of the horns, in which apertures of 3 in. \times 3 in. and $1\frac{1}{2}$ in. \times $1\frac{1}{2}$ in. were made. The wavelength was then measured for the various apertures used, and the effect of these on the wavelength is shown in Table 4. During these experiments it was possible to observe the efficiency of the metal lenses in reducing the beam divergence of the transmitting horn and in increasing the effective receiving area of the receiving horn. Removal of the

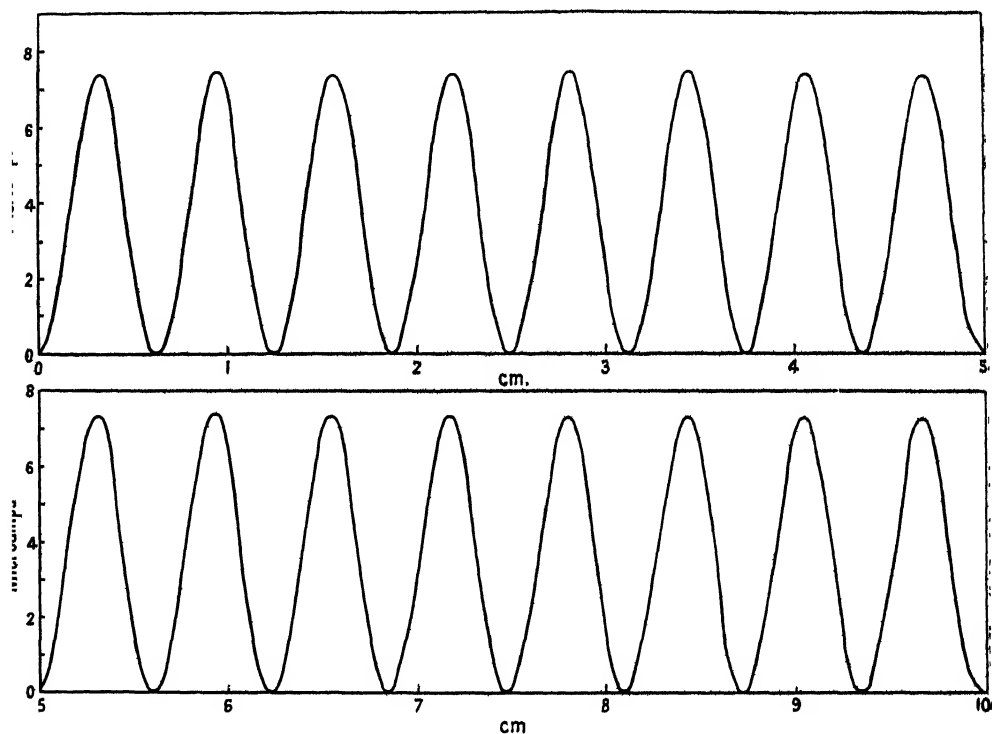


Figure 5. Plot of responses of the interferometer at position 1.

lens from the transmitting horn reduced the output at a maximum from 100 microamps to 32 microamps, with a similar reduction when the receiving horn lens was removed. Removal of both these lenses caused the maximum to fall from 100 microamps to 16 microamps.

Figure 7 shows the maximum and minimum responses obtained when the transmitting aperture is reduced with the interferometer in position 4; Figure 8 shows the variation in the amplitude of the maxima as M1 moves over its traverse of 50 cm. at positions 1 and 4, and the variation for the various apertures used at position 4. Figure 9 shows the effect of increasing the transmitted power on the sensitivity at the minimum.

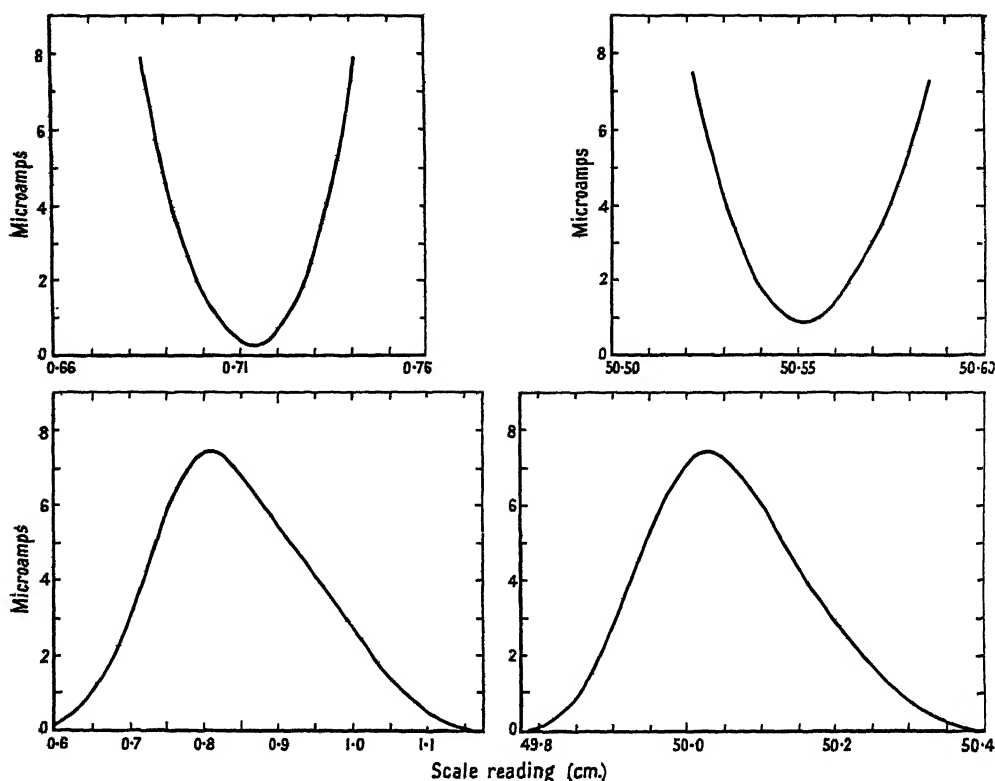


Figure 6. Maximum and minimum responses at position 4; 6 in. \times 6 in. apertures.

From Table 4 it appears that an accurate wavelength determination is possible at reduced spacings when the dimensions of the transmitting aperture are also reduced, this being probably due to the consequent shift of the measurements into the Fraunhofer region of diffraction. The responses are now, however, excessively asymmetrical (Figure 7), due to the increased beam divergence of the smaller transmitting apertures. This asymmetry is particularly marked in the case of the maximum responses where double humps occur. The reason for

Table 4. Effect of various Aperture Sizes on the Measured Wavelength at Position 4. Metal lenses used with the aperture unless otherwise stated.

Apertures		Wavelength (cm.)	
Transmitter	Receiver	Using minimum	Using response
6 in. \times 6 in.	6 in. \times 6 in.	1.2459	1.2459
3 in. \times 3 in.	6 in. \times 6 in.	1.2452	1.2452
1½ in. \times 1½ in.	6 in. \times 6 in.	1.2451	1.2448
6 in. \times 6 in.	3 in. \times 3 in.	1.2460	1.2461
3 in. \times 3 in.	3 in. \times 3 in.	1.2461	1.2461
6 in. \times 6 in.	1½ in. \times 1½ in.	1.2457	1.2457
6 in. \times 6 in.—lens off	6 in. \times 6 in.—lens off	1.2468	
6 in. \times 6 in.	6 in. \times 6 in.—lens off	1.2454	
6 in. \times 6 in.—lens off	6 in. \times 6 in.	1.2453	
3 in. \times 3 in.—lens off	6 in. \times 6 in.	1.2461	
1½ in. \times 1½ in.—lens off	6 in. \times 6 in.	1.2462	

these is not clear; they do not seem to be due to the secondary lobes in the polar diagram of the transmitting aperture; presumably they are due to reflection effects arising at the conducting plates used to reduce the transmitting aperture.

Use of the interferometer for the measurement of small frequency changes has also been made in position 5. The change in the position of a minimum due to a small frequency change depends on the relative distances of the two mirrors

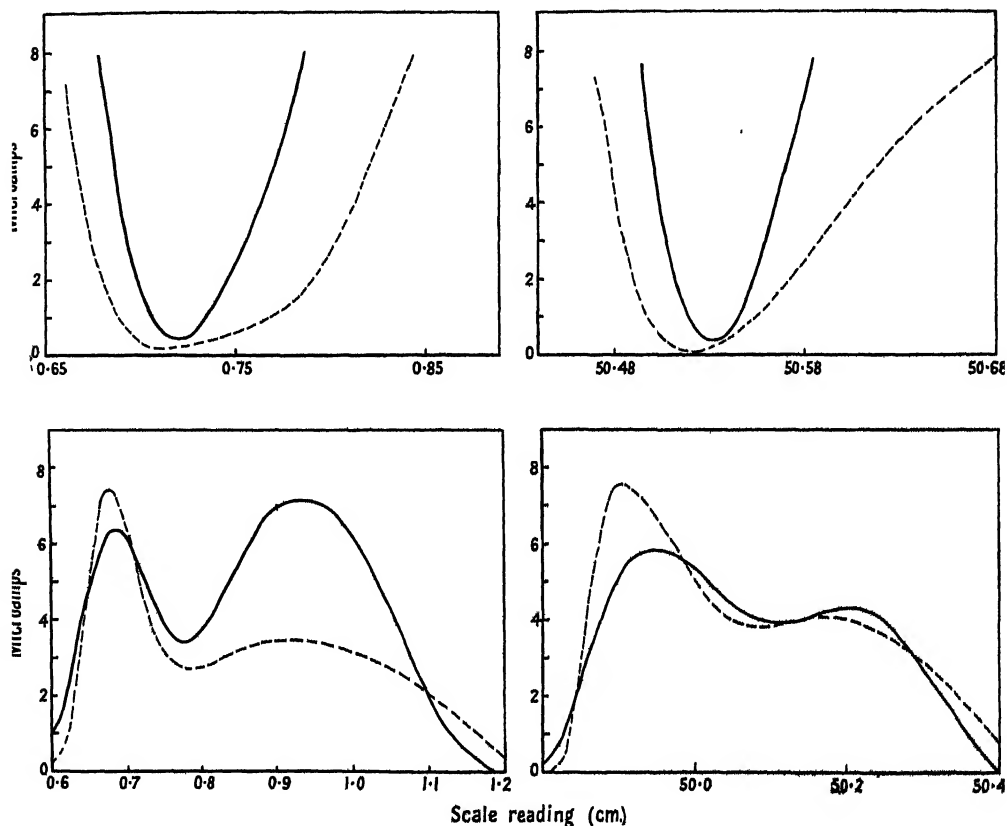


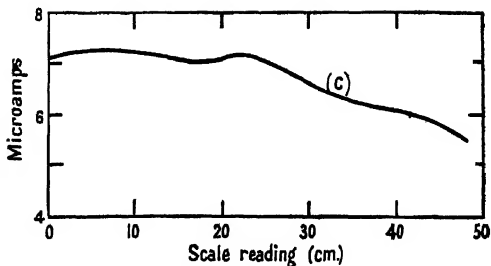
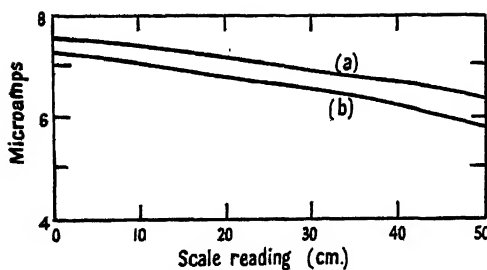
Figure 7. Maximum and minimum responses at position 4.

Curve ——— Transmitted aperture 3 in. \times 3 in.
 Receiving aperture 6 in. \times 6 in.
 Curve - - - - Transmitted aperture $1\frac{1}{2}$ in. \times $1\frac{1}{2}$ in.
 Receiving aperture 6 in. \times 6 in.

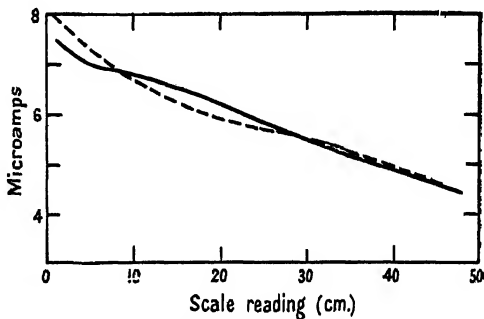
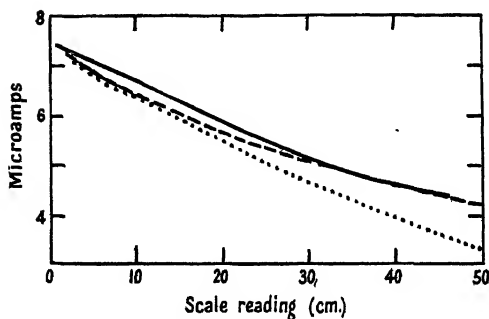
from the beam divider. Actual measured shifts with the interferometer in position 5 amounted to about 1 millimetre for a 10 Mc/s. change in operating frequency; this is in fair agreement with the theoretical value for this position assuming plane wave trains in the interferometer.

4.3. Dielectric Constant Measurements

Measurements of the dielectric constants of Perspex and polystyrene have been made in order to assess the value of the interferometer for this purpose. Two values of thickness of material were used, and measurements were made at all positions of the interferometer as well as with the various transmitting and receiving apertures used in position 4. The values quoted in Table 5 are those



$A_T=6$ in. \times 6 in. $A_R=6$ in. \times 6 in. (a) position 1; (b) position 4; (c) position 4;
both apertures without lenses.



Full line: $A_T=3$ in. \times 3 in. $A_R=6$ in. \times 6 in.
Dotted line: $A_T=3$ in. \times 3 in. $A_R=3$ in. \times 3 in.
Dashed line: $A_T=6$ in. \times 6 in. $A_R=3$ in. \times 3 in.
All curves refer to position 4.

Full line: $A_T=1\frac{1}{2}$ in. \times 1 in. $A_R=6$ in. \times 6 in.
Dashed line: $A_T=6$ in. \times 6 in. $A_R=1\frac{1}{2}$ in. \times 1 in.
All curves refer to position 4.

Figure 8. Variation of amplitude with transverse of mirror 1 at the various positions and apertures used in the interferometer. A_T =transmitting aperture; A_R =receiving aperture.

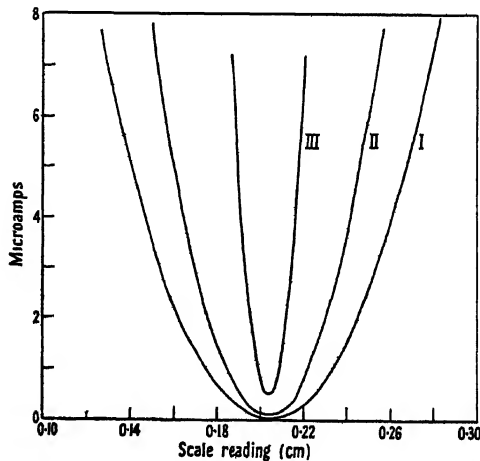


Figure 9. Response at a minimum for various current amplitudes.
Maximum current: Curve I, 70 μ a. Curve II, 150 μ a. Curve III, 720 μ a

obtained at position 1 using the 6 in. \times 6 in. apertures. The maximum deviation from these values over the full range of measurements amounted to about 1%, so that presumably the 'edge effect', due to the sheet being the same size as the mirror, is small.

Table 5. Dielectric Constants of Polystyrene and Perspex at 24,000 Mc/s. measured on the Interferometer

Polystyrene Sheet thickness	Dielectric constant		Perspex Sheet thickness	Dielectric constant	
	Approx.	Accurate		Approx.	Accurate
0.212 in.	2.43	2.63	0.183 in.	2.45	2.71
0.382 in.	2.66	2.63	0.252 in.	3.05	2.76

The values of the dielectric constants of Perspex and polystyrene obtained using the interferometer are consistent and are within a few per cent of values obtained otherwise.

No special precautions were taken in the preparation of these specimens, which were ordinary commercial sheets of the materials, and further experiments seem desirable in order to establish the accuracy of the method.

4.4. *Suggested Improvements*

For accurate measurements the responses of the maximum and minimum on the interferometer at the particular position used should be plotted. Ideally the symmetry of these responses, particularly that of the minimum, should approach the required limits of accuracy; in order to obtain this symmetry the use of larger apertures would seem advantageous.

One way of overcoming the effects of stray radiation and diffraction would be to enclose the interferometer in metal tubes of sectional dimensions about 6 in. \times 6 in. Such a system should enable higher sensitivities to be obtained. The chief disadvantages of such a system, however, are the possibility of generating higher order modes in such tubes, and the dependence of the wavelength on the dimensions of the tube. These effects could be investigated by operating the measuring mirror both in such a tube and in free space.

There are distinct possibilities of increasing the accuracy of the wavelength measurement, and hence of the determination of the velocity of electromagnetic waves, but it is clear that a more accurate and robust interferometer would be required, together with accurate control of temperature, pressure and humidity.

§ 5. USE AS A SUBSTANDARD OF LENGTH

The application of the interferometer for the accurate calibration of relatively large lengths seems possible. For this purpose adequate discrimination at the minimum, when large path differences exist in the interferometer, would be required.

From the results obtained in position 5, where the path difference was about 12 metres, it seems clear that the conditions for adequate discrimination, namely high-frequency stability of the source and beams of equal intensity, can be met.

In order to measure a length of 5 metres to one part in 10^6 the wavelength, as determined by the interferometer in the particular position used, would have to remain constant over this length to one part in 10^6 (cf. diffraction effects).

To verify this it would be necessary to increase the accuracy of the wavelength determination by a factor of 100. A better procedure would be to calibrate the interferometer in the particular position used against a standard of length, when a setting accuracy of 10^{-4} cm. would be required for lengths of 5 metres.

ACKNOWLEDGMENTS

The author would like gratefully to acknowledge the assistance given by Mr. J. N. Cruickshank in the design and operation of the Pound Frequency Stabilizer. The work described in this paper was carried out for the Department of Scientific and Industrial Research at the Telecommunications Research Establishment, Ministry of Supply. Acknowledgment is made to the Chief Scientist, Ministry of Supply, the Director of Radio Research, Department of Scientific and Industrial Research, and to the Controller, H.M. Stationery Office, for permission to publish the paper.

REFERENCES

- VON HIPPEL, A., and ROBERTS, S., 1946, *J. Appl. Phys.*, **17**, 610.
 KOCK, W. E., 1946, *Proc. Inst. Radio Engrs.*, **34**, 828.
 POUND, R. V., 1946, *Rev. Sci. Instrum.*, **17**, 490, *Proc. Inst. Radio Engrs.*, **35**, 1405.
 STRATTON, J. A., 1941, *Electromagnetic Theory* (London and New York: McGraw Hill Book Co.), § 8.13.
 WOOD, R. W., 1934, *Physical Optics* (New York: Macmillan Book Co.), p. 292.

Note on the Focusing of Electron Beams in certain Magnetic Fields

By P. A. STURROCK

National Bureau of Standards, Washington, D.C.

Communicated by L. Marton; MS. received 13th March 1950, and in amended form 3rd July 1950

ABSTRACT. Equations are set out which determine the focusing properties of electron beams in magnetic fields whose scalar potential has a plane of antisymmetry. From these is derived the condition that a proposed ray-axis and associated focusing requirements should be physically realizable. It is also shown that the fringe-effect of fields with sharply defined boundaries may be characterized by a pair of focal lengths for which formulae are given.

§ 1. INTRODUCTION

MAGNETIC fields whose scalar potential possesses a plane of antisymmetry, the 'equatorial' plane, have long been used in mass spectrometers and beta-ray spectrometers (Siday 1947). Such fields are often prescribed to be constant over the plane within defined boundaries; ray-tracing in the equatorial plane is then simplified; moreover, such fields may be approximately realized by magnets having flat-faced pole pieces. The fringe effect is usually treated as a correction (Coggeshall 1947).

§ 2. THE PARAXIAL EQUATIONS

Consider a magnetic field whose scalar potential has a plane of antisymmetry; the field is then determined by the (normal) field strength H in the equatorial plane. Let a particular ray in the equatorial plane be the 'ray-axis'. We may

then give to a point P coordinates (u, v, w) , where $w = PM$, M being the foot of the normal from P to the equatorial plane, $v = MN$, N being the foot of the normal from M to the ray-axis, and u measures the arc-length of N along the ray-axis (Figure 1).

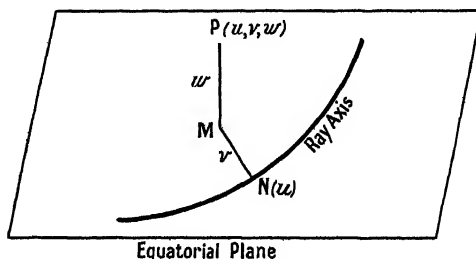


Figure 1.

If $\sigma(u)$ is the curvature of the ray-axis, measured so as to be positive when the centre of curvature lies on the side of the ray-axis for which v is positive, then

$$\sigma = H^0/p, \quad \dots\dots(1)$$

where $H^0(u)$ is the field strength on the ray-axis, measured in the direction of the w -axis, and p is the $H\rho$ momentum of the beam.

Rays within a certain neighbourhood of the ray-axis may be characterized by equations between their coordinates of the form $v = v(u)$ and $w = w(u)$. Then for rays indefinitely close to the ray-axis, the functions $v(u)$ and $w(u)$ are solutions of the following second-order differential equations:

$$\frac{d^2v}{du^2} + Vv = 0, \quad \text{wherein } V = (H^0/p)^2 - H^0_v/p, \quad \dots\dots(2)$$

and

$$\frac{d^2w}{du^2} + Ww = 0, \quad \text{wherein } W = H^0_v/p, \quad \dots\dots(3)$$

the function $H^0_v(u)$ being the gradient of H in the v direction, evaluated on the ray-axis.

The equations (2) and (3) are formally identical with the equations determining the paraxial behaviour of rays in an optical system with two orthogonal planes of symmetry. It is seen at once that if H^0_v is zero, as in a uniform field, then V is positive and W is zero, so that there is convergent focusing of rays in the equatorial plane but no focusing of rays diverging from the plane.

If the field is confined to a small region along the ray-axis, we may deduce from equations (2) and (3) that the focusing action of the field may be represented by a pair of focal-lengths, f_v and f_w , where

$$1/f_v = \int V du \quad \text{and} \quad 1/f_w = \int W du, \quad \dots\dots(4)$$

the integration extending over the field.

§ 3. CONDITION FOR A PROPOSED BEAM TO BE PHYSICALLY REALIZABLE

The necessary and sufficient condition that a ray-axis, characterized by a function $\sigma(u)$, and an associated assembly of the neighbouring rays, characterized by functions $V(u)$ and $W(u)$, should be physically realizable is that

$$V + W = \sigma^2. \quad \dots\dots(5)$$

Examination of equations (1), (2) and (3) shows that this is necessary; it is also sufficient, for, if the condition is satisfied, values of $H^0(u)$ and $H^0_v(u)$ may be determined which will realize the rays.

It follows at once that, for a short field,

$$1/f_v + 1/f_w = \int \sigma^2 du, \quad \dots\dots(6)$$

the integral extending over the field.

From either of equations (5) and (6) it is possible to derive the behaviour of rays off the equatorial plane from knowledge of their behaviour in the plane: (5) may be treated as a formula for W , (6) as a formula for f_w .

A simple example of this corollary would be to find a realizable ray-assembly which achieves concurrent v - and w -focusing. This is certainly obtained if $V=W$ for all u : suppose that, in addition, each function is constant. If $V=W=1/2a^2$, equation (5) asserts that the rays will be realizable if, and only if, $\sigma=1/a$. We see that the ray-axis is a circle of radius a and that any two points on the circumference subtending an angle $\pi\sqrt{2}$ to the centre are 'doubly' conjugate. Equations (1) and (3) show that H^0 and H^0_v should have the values p/a and $p/2a^2$, respectively. The system is recognized to be the 'double-focusing' beta-ray spectrometer of Svartholm and Siegbahn (1947).

§4. THE FRINGE-EFFECT

Suppose that H has the value H^* over a certain domain of the equatorial plane, outside of which it falls rapidly to zero. The 'fringe-effect' is the effect of the field outside the domain boundary (which, in practice, will approximate to the contours of the pole pieces). It is assumed that the extent of the fringe field is short compared with the radius of curvature of the ray-axis.

Suppose that the ray-axis intersects the field-boundary at O (Figure 2), making an angle θ with the normal at O directed inwards. If H_n is the field derivative along the normal, we see that, in the neighbourhood of O ,

$$H^0_v = H_n \sin \theta. \quad \dots\dots(7)$$

Moreover, an integral along the ray-axis may be transformed into an integral along the normal by the relation

$$dn = du \cos \theta. \quad \dots\dots(8)$$

It is clear from equations (1) and (8) that, if ψ represents the deflection of the ray-axis by the fringe-field,

$$\psi = \sec \theta \int_{fr} (H/p) dn. \quad \dots\dots(9)$$

The earlier assumption about the extent of the fringe-field implies that ψ is small.

The focusing action of the fringe-field may be evaluated by equations (4) together with equations (7) and (8). We find that the effect of the fringe field may be characterized by two focal-lengths, given by the formulae

$$1/f_v = \sec \theta \int_{fr} (H/p)^2 dn - \tan \theta (H^*/p), \quad \dots\dots(10)$$

and

$$1/f_w = \tan \theta (H^*/p). \quad \dots\dots(11)$$

Of the two terms making up formula (10), the former represents the effect of the finite extension of the fringe-field, the latter the effect of the inclination of the beam to the field-boundary; only the latter effect appears in formula (11). The

latter term of formula (10) is taken into account in ray-tracing, so that it is the former term only which is usually interpreted as 'fringe-effect'. Since, for a narrow fringe, this term is small, the most important fringe-effect is seen to be

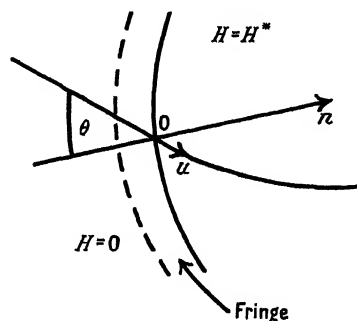


Figure 2.

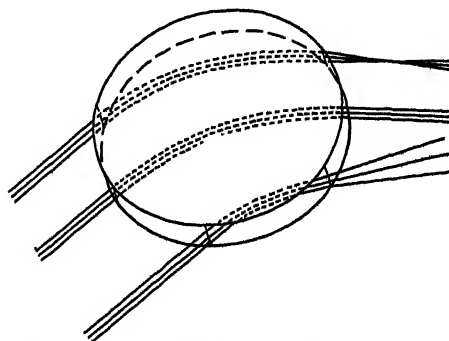


Figure 3. Focusing in w -direction due to the fringe-field of circular pole pieces.

that represented by formula (11). Figure 3 demonstrates how the fringe-focusing in the w -direction depends on the angle of intersection of the beam and the field-boundary.

§ 5. CONCLUSION

It is hoped that this note will prove useful in visualizing the focusing of the kind of fields discussed, and in picturing the behaviour of beams in fringe-fields. Neglect of the w -focusing of the fringe-field of a beta-ray spectrometer will, in general, result in heavy loss of beam current; by taking this focusing into account, transmission would be substantially improved.

These results are extracted from an analysis of the image-forming and image-marring properties of arbitrary magnetic fields, which it is hoped to publish at a later date. It has been pointed out to the author that the substance of the note is also implicit in the work of Cotte (1938).

REFERENCES

- COGGESHALL, N. D., 1947, *J. Appl. Phys.*, **18**, 855.
 COTTE, M., 1938, *Ann. Phys., Paris*, **10**, 333.
 SIDAY, R. E., 1947, *Proc. Phys. Soc.*, **59**, 905.
 SVARTHOLM, N., and SIEGBAHN, K., 1947, *Ark. Mat. Astr. Fys.*, **33A**, No. 21.

Satellite Resonances in Ultrasonic Interferometry

By J. F. W. BELL

Kings' College, Newcastle-upon-Tyne

MS. received 14th February 1950

ABSTRACT. An experimental investigation into the origin of satellite resonances has been carried out by the author. They are identified as mode resonances of the gas in the interferometer tube and are of the type described by Lord Rayleigh.

The presence of unresolved satellites in the principal interferometer resonance introduces a considerable error into absorption measurements. Results obtained by Van Itterbeek and his co-workers are shown to be in agreement with the Krasnooshkin interferometer theory, which takes into account the effect of the multiple nature of the principal resonance.

A criterion for the choice of crystals for ultrasonic absorption measurements is given.

PART I.—THE ORIGIN OF SATELLITE RESONANCES

§ 1. INTRODUCTION

AN effect frequently encountered in the use of the ultrasonic interferometer for measurements on gases is the occurrence of more than one series of resonances. The extra resonances fall into two categories, namely those which are due to lack of parallelism between crystal and reflector, and the rest which are characteristic of the crystal and interferometer tube used. The former are eliminated when crystal and reflector are exactly parallel.

The latter, which have been widely discussed in the literature, have been the subject of the experimental investigation described in this paper. The original description of these resonances, which, because of their proximity to the principal resonance, are known as satellites, was given by Pielemeier (1931). He deduced that the presence of more than one series of resonances indicated that the ultrasonic energy was being propagated with more than one velocity. This he classified with the variation of velocity with intensity observed near explosions.

It was soon evident that this explanation could not be correct as the character of the resonances did not change with the power supplied to the generator.

In a second paper Pielemeier (1938) pointed out this error in his previous work and suggested that more than one frequency was present in the vibrating crystal. These other frequencies gave rise to differently spaced resonances. It is very evident that such an explanation is inapplicable to the case where the crystal is made to resonate by impressing a voltage. In the steady state the crystal will vibrate at the applied frequency only.

§ 2. EXPERIMENTAL INVESTIGATION

Preliminary measurements on air were carried out with a $1\frac{1}{2}$ -in. diameter cylindrical interferometer using a 1-in. diameter X-cut quartz crystal operating at a frequency of 250 kc/s. A radio-frequency bridge was developed to measure the resistive component of the equivalent electrical circuit of the crystal. A part of this resistance arises from the acoustic radiation from the vibrating faces of

the crystal. Figure 1 shows the variation of electrical resistance of the crystal with the length of the column of air between the crystal and reflector.

At resonances, which occur every half-wavelength, the coupling between the crystal and the column of air increases. This results in an increase in the equivalent electrical resistance of the crystal.

It will be observed that the length of air column is slightly greater for satellites than for the corresponding principal resonance. The separation increases as the order of resonance increases. On plotting the distances of the satellites from the principal against the length of the column corresponding to the principal resonance, a series of straight lines, all passing through the origin, is obtained (Figure 2). Thus a wavelength λ_n can be ascribed to each satellite. The slopes of the lines give the fractional excess of satellite wavelength over principal wavelength λ .

To discover the source of a phenomenon such as this, an experimental approach often used is to vary the conditions and instrument parameters as far as possible. If circumstances can be found which modify the phenomenon in

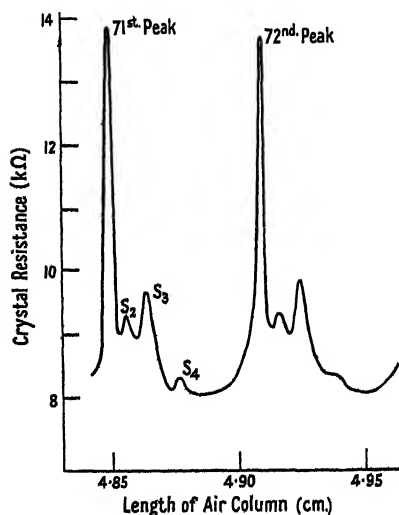


Figure 1. Satellite patterns for air.

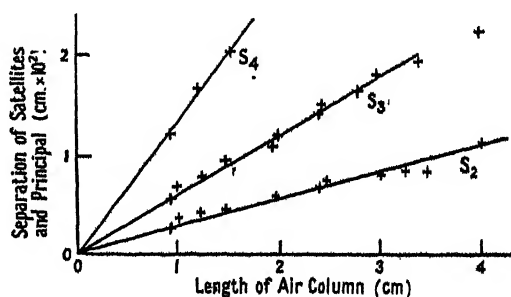


Figure 2. Plots of the distances of the satellites from the principal resonance as the length of the column of air is varied.

any way, a valuable clue to its origin is obtained'; at the least this will give rise to the deduction of a positive experimental law of behaviour.

The first change attempted was to line the walls of the interferometer tube with a layer of cotton wool in order to damp out any reflection from the side walls. The results were inconclusive in that all resonances were reduced in strength, but no change in wavelength could be detected.

It was next decided to investigate the effect of change of principal wavelength. Selection of gases giving a wide range of velocities of sound, together with low absorption (to give good resolving power) was necessary. Hydrogen has by far the greatest velocity of sound of the gases readily obtainable; in addition, its absorption coefficient is small. For velocities intermediate between air and hydrogen, mixtures of air, hydrogen and coal gas were convenient. For velocities less than that of air no gas considered was suitable as, although the velocity in polyatomic gases is low, the absorption, because of molecular effects, is high.

As it was found that the satellites were very close to the principal at the shorter wavelengths, no observations on low velocity gases were attempted.

Figure 3 shows a series of observations for different gases and gas mixtures.

§ 3. THEORETICAL CONSIDERATIONS

In a recent paper by Krasnooshkin (1944) on the theory of the interferometer it was pointed out that the non-uniform vibration of the crystal surface could result in the excitation of transverse modes of resonance in an interferometer tube. An excellent account of these modes at audio frequencies, first discovered

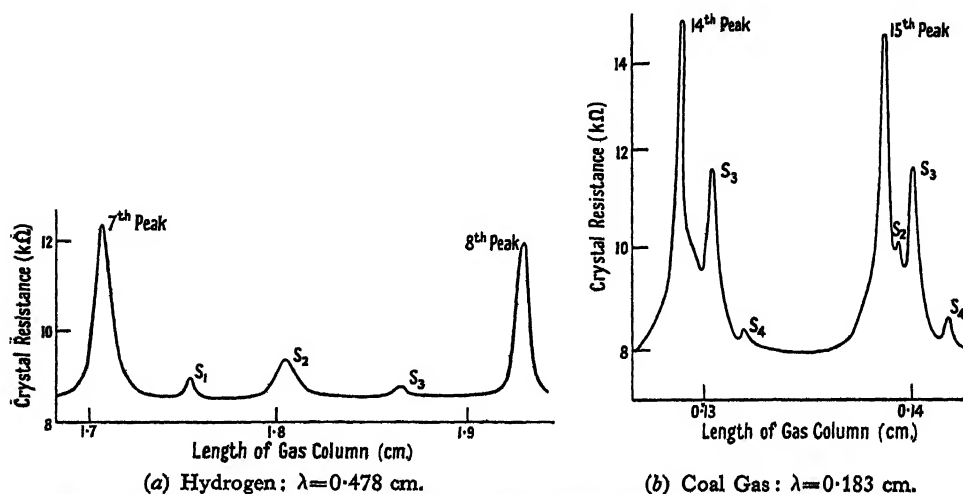


Figure 3. Plots of the satellite resonances for the two gases giving widely different wavelengths.

theoretically by Lord Rayleigh, is given by Hartig and Swanson (1938). Their electromagnetic counterparts are well known to workers on centimetre radio waves.

A characteristic property is the dispersion of their velocity of propagation. Thus

$$\lambda_s = \lambda / \left\{ 1 - \left(\frac{\lambda}{\lambda_0} \right)^2 \right\}^{1/2}. \quad \dots (1)$$

Here λ is the wavelength for plane waves (the principal wavelength is assumed to be equal to λ) and λ_0 is a function of the tube diameter d , which was equal to 3.8 cm., and the mode :

$$\lambda_0 = \pi d / x_{MN}. \quad \dots (2)$$

x_{MN} is the M th root of the first derivative of the Bessel function of the first kind of order N . For complex modes x_{MN} is large and λ_0 is small. When $\lambda = \lambda_0$, λ_s is infinite; at this wavelength the phase velocity is infinite and energy can no longer be propagated in the mode; λ_0 is known as the cut-off wavelength for the particular mode.

In the experimental case investigated $\lambda \ll \lambda_0$. Thus equation (1) approximates to

$$\frac{\lambda_s - \lambda}{\lambda} = \frac{1}{2\lambda_0^2} \lambda^2. \quad \dots (3)$$

If satellites are mode resonances their wavelengths must follow this law.

§ 4. INTERPRETATION OF RESULTS

Figure 4 shows $(\lambda_s - \lambda)/\lambda$ plotted against λ^2 . The straight lines obtained show that equation (3) is followed quite well. It only remains to see whether reasonable values of α_{MN} are obtained from these observations.

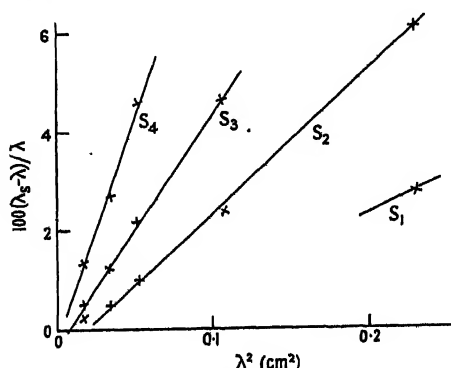


Figure 4. The experimental test of equation (3). The slopes of these lines give the cut-off wavelengths of the modes of vibration.

Table 1 shows values of λ_0 and $\pi d/\lambda_0$ obtained. In every case a value of α_{MN} can be found to correspond to $\pi d/\lambda_0$; the values of M and N are also shown.

Table 1

Satellite	S_1	S_2	S_3	S_4
λ_0 (cm.)	1.85	1.32	1.04	0.91
$\pi d/\lambda_0$	6.4	9.0	11.4	13
α_{MN}	6.7	8.5	11.7	13.3
N	2	1	1	0
M	2	3	4	4

λ_0 is the cut-off wavelength for the various satellites shown in Figure 3. $\pi d/\lambda_0$ is the experimental value of α_{MN} .

There thus remains little doubt as to the identification of satellite resonances with Rayleigh modes.

The lowest order modes have not been detected, presumably because they are so close to the principal resonance that they are not resolvable. The measurements on hydrogen confirm this opinion, for a satellite is observed which was unresolvable in the case of air.

The compounding of the principal resonance with low order modes gives a possible explanation of the fact that the lines of Figure 4 do not pass quite through the origin. The wavelength of the principal resonance has been taken as λ as the wave front was assumed to be plane. If it is actually composed of a number of modes, each having a wavelength greater than λ , the observed wavelength will be slightly in excess of λ . Thus the observed values of $(\lambda_s - \lambda)/\lambda$ will be less than the true values. This would result in a general raising of the lines of Figure 4.

PART II.—THE INFLUENCE OF UNRESOLVED SATELLITES ON ABSORPTION MEASUREMENTS

§ 5. THEORY

Krasnooshkin (1944) has taken into account the fact that the principal resonance peak may consist of a number of low order modes rather than the simple plane waves assumed hitherto. He showed that under these conditions

the observed absorption α_{obs} is made up of two factors : (i) the true absorption of acoustic energy by the modes making up the principal resonance—this is slightly greater than the value for plane waves ; (ii) the interferential absorption α_I which is due to interference effects between the various modes: this effect is expected to be quite large.

$$\alpha = \frac{\alpha_{\text{obs}} - \alpha_I}{1 + \chi\lambda^2/4\pi^2} \quad \dots\dots(4)$$

$$\alpha_I = \frac{\Delta\chi\lambda}{16\pi} \quad \dots\dots(5)$$

α is the absorption coefficient per centimetre for plane waves in the gas, λ is the free space wavelength, χ and $\Delta\chi$ are constants of the crystal and interferometer of dimensions cm^{-2} . χ expresses the departure of the velocity of the combined modes from that of plane waves and $\Delta\chi$ expresses the multiple nature of the resonance; for a plane wave $\chi = \Delta\chi = 0$; for any pure mode $\Delta\chi = 0$.

The $\chi\lambda^2/4\pi^2$ term in the denominator of equation (4) corrects the absorption of the complex mode system to that of plane waves. As it amounts to only a few per cent it will be neglected.

§ 6. EXPERIMENTAL TESTS

It has recently been pointed out by Kittel (1946-7) that the measurement of low absorption coefficients—those of monatomic and diatomic gases—is very unsatisfactory. There is a lack of correlation between values obtained at different frequencies and by different observers. In addition the values are in excess of the formula of Stokes and Kirchhoff giving the absorption due to viscosity and thermal conductivity.

The Krasnooshkin theory at once suggests the reason for this: what is actually being measured is the sum of α and α_I ; as the value of α_I will be different for each crystal used the divergence obtained is not surprising.

A method of separating the two absorptions is available as α is inversely proportional to the pressure. Expressing equation (4) in the form necessary for the determination of α_I

$$\alpha_{\text{obs}} = \frac{\alpha_0}{p/p_0} + \alpha_I \quad \dots\dots(6)$$

where p is the pressure and α_0 the gas absorption corresponding to atmospheric pressure p_0 . A plot of α_{obs} against $1/p$ will give a straight line of slope α_0/p_0 and intercept α_I (see Figure 5). If, using the same crystal and interferometer, a determination of α_I is made for gases of different velocities equation (2) can be tested.

Measurements suitable for this test have been carried out by Van Itterbeek and his co-workers (Van Itterbeek and Mariens 1937, Van Itterbeek and Thys 1938) using a 304.4-kc/s. crystal. The use of oxygen and hydrogen at a number of temperatures gave a wide range of velocities.

Table 2 summarizes the results deduced from these data, and graphs of (α_{obs} , $1/p$) are shown in Figure 5.

The accuracies of values of α_0 given in the last column of Table 2 are calculated on the assumption that the measured values α_{obs} were accurate to $\pm 5\%$.

Plotting α_I against λ (Figure 6) gives a straight line passing through the origin, showing that equation (5) is followed. The value of $\Delta\chi$ deduced was 35 cm^{-2} .

§ 7. CONCLUSION

The experimental data available are in good agreement with Krasnooshkin's theory. The values of absorption are of the same order of magnitude as those calculated from the formula of Stokes and Kirchhoff, but the accuracy leaves

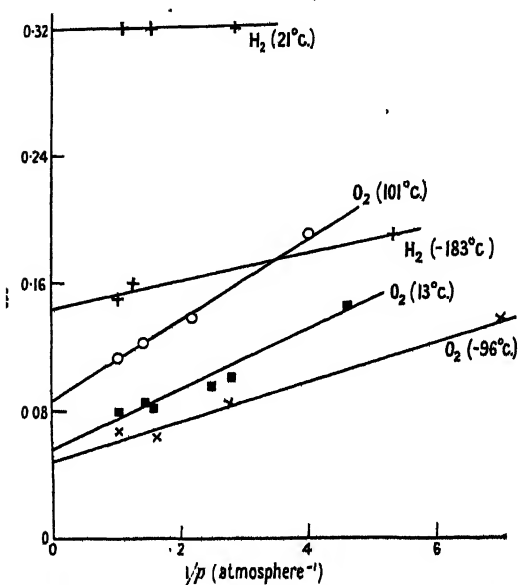


Figure 5. The graphs show the variation of measured absorption coefficient with pressure for gases of various velocities.

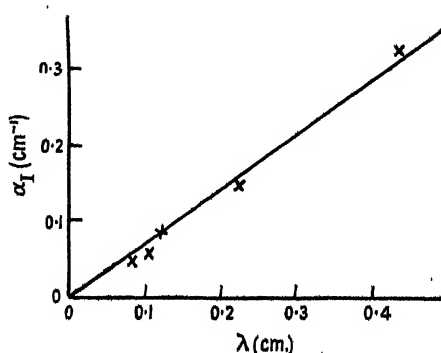


Figure 6. α_I is the absorption coefficient that would be observed if a gas of zero absorption coefficient and wavelength λ were used in the interferometer.

much to be desired because of the masking effect of a large value of α_I . Crystals with a simple vibration of the radiating surface, the ideal being piston-like motion, give the low values of α_I desirable. A criterion for the selection of crystals as

Table 2. Interpretation of the Results of Van Itterbeek's Measurements in Terms of Krasnooshkin's Theory

Gas	Temperature (° c.)	λ (cm.)	α_I (cm ⁻¹)	α_0 (cm ⁻¹)	α_0 (calc.) (cm ⁻¹)	Estimated error in α_0 (cm ⁻¹)
Oxygen	110	0.122	0.084	0.026	0.032	± 0.005
Oxygen	13	0.107	0.057	0.017	0.032	± 0.004
Oxygen	— 96	0.083	0.049	0.013	0.025	± 0.003
Hydrogen	12	0.436	0.32	0.000	0.015	± 0.015
Hydrogen	— 183	0.226	0.145	0.008	—	± 0.007

α_0 (calc.) gives the values calculated from the formula of Stokes and Kirchhoff.

ultrasonic generations is thus available. Considerable work on the measurement of $\Delta\chi$ for crystals of different sizes and frequencies, leading to an accurate determination of the absorption coefficients of gases of the type mentioned above is still to be carried out.

ACKNOWLEDGMENTS

The work was carried out at University College, Hull. The author wishes to thank Professor L. S. Palmer for placing the research facilities of the Physics Department at his disposal, Mr. J. M. Hough for his discussions and assistance and Mr. Taylor for construction of the interferometer.

REFERENCES

- HARTIG, H. E., and SWANSON, A. W., 1938, *Phys. Rev.*, **54**, 618.
 KITTEL, C., 1946-7, *Rep. Prog. Phys.*, **11**, 216 (London: Physical Society).
 KRASNOUSHKIN, P. E., 1944, *Phys. Rev.*, **65**, 190.
 RAYLEIGH, Lord, 1900, *Theory of Sound*, Vol. II (London: Macmillan and Co.), p. 279.
 PIELEMEIER, W. H., 1931, *Phys. Rev.*, **37**, 1682, **38**, 1236; 1938, *J. Acoust. Soc. Amer.*, **9**, 212.
 VAN IITERBEEK, A., and MARIENS, P., 1937, *Physica*, **4**, 609.
 VAN IITERBEEK, A., and THYS, L., 1938, *Physica*, **5**, 889.

LETTERS TO THE EDITOR

Frictional Relaxation Oscillations

In their paper on static friction, Parker and Hatch (1950) seek to explain the so-called 'stick-slip' phenomenon—encountered when frictional force is measured between an elastically restrained body and a driven surface, and now generally agreed to be the exhibition of relaxation oscillations by the measuring instrument and not due to the intrinsic mechanism by which two bodies in frictional contact slide relative to each other—and in doing so assume that, during some part of the cycle, the two surfaces are at rest relative to each other, and static friction consequently comes into operation. This assumption is not generally true. Indeed, generally during the frictional generated relaxation oscillation cycle, the two surfaces are never at rest relative to each other. This is readily apparent from the fact that during the 'stick' part of the cycle the dragged elastically restrained surface never catches up with the dragging surface, always having a velocity less than it (Bristow 1942, 1947).

This erroneous assumption is all too commonly made, and has been exaggerated by the use of the apparently realistic, but actually unrealistic and misleading, term 'stick-slip'.

Dudley and Swift (1949) have recently given an elegant analysis of the motion of elastic restrained body subject to frictional traction by contact with a second body in a large variety of friction-velocity relationships and, attempting to explain the writer's observations, only deduce the exhibition of relaxation oscillations if a period of rest is encountered during which static friction becomes operative.

Neither of these two papers therefore gives satisfactory explanation of frictional relaxation oscillation and consequently gives no clue as to the nature and mechanism of either 'dry' or 'boundary' kinetic friction.

The Motor Industry Research Association,
 Great West Road,
 Brentford, Middlesex.
 19th July 1950.

J. R. BRISTOW.

- BRISTOW, J. R., 1942, *Nature, Lond.*, **149**, 169; 1947, *Proc. Roy. Soc. A*, **189**, 88.
 DUDLEY, B. R., and SWIFT, H. W., 1949, *Phil. Mag.*, **40**, 849.
 PARKER, R. C., and HATCH, D., 1950, *Proc. Phys. Soc. B*, **63**, 185.

It is clear from Dr. Bristow's letter that our explanation of the so-called 'stick-slip' phenomenon is not sufficiently full. The behaviour of a soft metal hemisphere in contact with a hardened flat non-metal surface, under the influence of increasing frictional force, will therefore be stated in more detail.

Figure 10 of our paper shows that relative motion takes place between the two surfaces at extremely small tangential loads, and that the total amount of movement is dependent upon

the magnitude of the tangential load. This movement is accompanied by an increase in the real area of contact, and when this area cannot grow further, any additional increase in the tangential force gives rise to slip in the 'macro' sense.

If, now, we imagine these two members to form a 'stick-slip' pair, that is the soft metal hemisphere to be elastically constrained and the glass member to move at a low uniform velocity, it will be seen that, at the moment of movement, the two surfaces will appear to move together for only a very short time, after which the relative velocity will gradually increase up to the point of macroslip. If it is now assumed that, during the period of macroslip, the two surfaces will become disturbed and present partly or completely new contact areas; this will, in turn, mean that the process will be repeated indefinitely. In this interpretation, it is therefore not necessary to assume that the two bodies are relatively at rest while the tangential force is increasing, that is while movement of the flat is taking place.

Messrs. Ferodo Ltd.,
Chapel-en-le-Frith,
Stockport, Cheshire.
23rd August 1950.

R. C. PARKER.
D. HATCH.

CONTENTS FOR SECTION A

	PAGE
Prof. P. B. MOON. The Hard Components of Scattered Gamma-Rays	1189
Mr. A. STORRUSTE. The Rayleigh Scattering of 0.41 mev. Gamma-Rays at Various Angles	1197
Prof. S. DEVONS and Mr. G. R. LINDSEY. γ -Radiation from the Resonant Capture of Protons by ${}^7\text{Li}$ Nuclei	1202
Dr. W. J. SWIATECKI. The Density Distribution inside Nuclei and Nuclear Shell Structure	1208
Mr. J. HUGHES and Dr. K. J. LE COUTEUR. Spin Orbit Coupling in the Nuclear Shell Model	1219
Dr. K. J. LE COUTEUR and Mr. S. ZIENAU. Coherent Scattering of Light by an Atom and Negative Energy States	1223
Dr. J. R. GREENING. The Determination of X-Ray Wavelength Distributions from Absorption Data	1227
Dr. S. L. ALTMANN. Relation between the Franck-Condon Frequencies of Absorption and Fluorescence for some Unsaturated Hydrocarbons	1234
Prof. C. B. A. MCCUSKER. Penetrating Particles in Air Showers	1240
Dr. E. P. GEORGE and Mr. J. EVANS. Observations of Cosmic-Ray Events in Nuclear Emulsions Exposed below Ground	1248
Prof. L. F. BATES and Dr. J. H. DAVIS. Heat Changes Accompanying Magnetization in Low and Moderate Fields: the Effects of Strain, and a Theoretical Interpretation	1265
Mr. C. BULL and Dr. G. F. J. GARLICK. The Luminescence of Diamonds	1283
Letters to the Editor:	
Dr. W. E. MOFFITT. The Ultra-Violet Spectrum of Ethylene	1292
Mr. R. A. DURIE. The Spectra of Flames Supported by Fluorine	1292
Dr. J. B. BIRKS. Scintillation Efficiency of Anthracene Crystals	1294
Mr. E. E. SALPETER. Dissociation Cross Sections for Fast Hydrogen Molecule Ions	1295
Dr. E. W. TITTERTON. The Reaction ${}^7\text{Li} \gamma p {}^6\text{He}$	1297
Mr. P. J. GRANT. Forbidden β -Decay in ${}^{24}\text{Na}$	1298
Reviews of Books	1300
Contents for Section B	1301
Abstracts for Section B	1301

ABSTRACTS FOR SECTION A

The Hard Components of Scattered Gamma-Rays, by P. B. MOON.

ABSTRACT. A summary is given of theoretical information on the three processes by which gamma-rays can be elastically scattered by atoms, viz. Rayleigh scattering by bound electrons, Thomson scattering by the nuclear charge and (exceptionally) nuclear resonant scattering.

Interference between the three scattered waves is considered and shown to be of practical importance as between Rayleigh and Thomson scattering of hard gamma-rays at large angles.

The calculated intensities of Rayleigh, Thomson and Compton scattering are plotted against angle of scattering for gamma-ray energies of 2.8 and 0.41 mev., and for scatterers of Al, Cu and Pb. The graphs illustrate the dominance of Rayleigh scattering at very small angles for all energies, and the comparable intensities of Rayleigh and Thomson scattering at high energies and large angles.

The unidentified hard component found by Pollard and Alburger in the large-angle scattering of 2.8-mev. gamma-rays by various elements is interpreted as a mixture of Rayleigh and Thomson scattering; theoretical and experimental intensities show reasonable agreement for both light (Al) and heavy (Pb) scatterers.

Experiments on the scattering of 0.41-mev. gamma-rays at about 115° are reported. In agreement with theory, about 2% of the radiation scattered from lead is found to retain the full energy; for copper and aluminium, the proportion of elastically scattered photons is much smaller.

The Rayleigh Scattering of 0.41 MeV. Gamma-Rays at Various Angles, by A. STORRUSTE.

ABSTRACT. A report is given on some measurements of the Rayleigh cross section for scattering of 0.41 mev. gamma-rays by lead, copper and aluminium. At small angles the Rayleigh cross section in lead is found to decrease rapidly with increasing angle of scattering from approximately 12 times the Compton cross section at 3° to a value equal to the Compton cross section at 12° . In copper the Rayleigh cross section is found to equal the Compton cross section at 4.5° . In aluminium no substantial excess scattering above the Compton cross section is found, the smallest angle at which measurements were taken being 3° . At 60° , 90° , 120° and 150° the hard component of the scattered radiation from lead is found to be respectively 1.77, 1.14, 0.88 and 0.66% of the Compton component as calculated by the Klein-Nishina formula.

γ -Radiation from the Resonant Capture of Protons by ${}^7\text{Li}$ Nuclei, by S. DEVONS and G. R. LINDSEY.

ABSTRACT. Further measurements have been made of the angular distribution of the γ -radiation in the reaction ${}^7\text{Li}(p,\gamma){}^8\text{Be}$. The results indicate approximate isotropy, at resonance, for both spectral components (17.6 and 14.8 mev.) of the radiations. The difficulties in reconciling the properties of the γ -radiation with results of recent experiments on the scattering of protons by lithium are discussed.

The Density Distribution inside Nuclei and Nuclear Shell Structure, by W. J. SWIATECKI.

ABSTRACT. A method of evaluating the density distribution of particles inside a heavy nucleus is described. The results are used to show that the electrostatic repulsion between protons is too weak to produce a nucleus sufficiently hollow to explain the observed 'closed shells' at N , $Z=50, 82$.

Spin Orbit Coupling in the Nuclear Shell Model, by J. HUGHES and K. J. LE COUTEUR.

ABSTRACT. It is shown that the magnitude of the doublet splitting postulated in Mayer's nuclear shell model is consistent with that observed in ${}^4\text{He}$. The strength of the indicated spin orbital interaction between pairs of nucleons is worked out.

Coherent Scattering of Light by an Atom and Negative Energy States, by K. J. LE COUTEUR and S. ZIENAU.

ABSTRACT. It is proved that the difference between the single electron theory and the positron theory for the coherent scattering of light by an atom represents the scattering of light by the atomic potential. In the practical cases the effect is negligible. The gauge invariance of the perturbation formulae for this effect is discussed.

The Determination of X-ray Wavelength Distributions from Absorption Data, by J. R. GREENING.

ABSTRACT. An existing empirical method which may be used, in limited cases, to determine x-ray wavelength distributions from absorption data is given some theoretical foundation. It is shown that the absorption curve function and the wavelength distribution function have the same relationship as a Laplace pair, the absorption curve function being the Laplace transform of the wavelength distribution function. The theory leads to the suggestion of Laplace pairs other than the one previously suggested, thus widening the scope of the method. By using the additive property of Laplace transforms the absorption method of determining spectral distributions is made of general application. The experimental requirements are discussed and examples of wavelength distributions determined by the absorption method are given.

Relation between the Franck-Condon Frequencies of Absorption and Fluorescence for some Unsaturated Hydrocarbons, by S. L. ALTMANN.

ABSTRACT. Differences between the Franck-Condon frequencies for emission and absorption are calculated for butadiene, naphthalene and anthracene. The molecular-orbital method is used and its validity in the present case is discussed. A correlation with the experimental data has been attempted for anthracene and appears to be satisfactory. It is shown that the calculated values may be correlated qualitatively with the fluorescence efficiencies of the molecules.

Penetrating Particles in Air Showers, by C. B. A. McCUSKER.

ABSTRACT. The penetrating particles in extensive air showers are compared directly with the penetrating particles in the main cosmic-ray beam. It is found that not more than $70 \pm 3\%$ of the air-shower particles are single μ -mesons, whilst the remainder are strongly interacting particles. The barometer coefficient of extensive penetrating showers is determined.

Observations of Cosmic-Ray Events in Nuclear Emulsions Exposed below Ground, by E. P. GEORGE and J. EVANS.

ABSTRACT. Ilford Nuclear research plates were manufactured in a laboratory at an equivalent depth of 60 m. of water below ground. After having been stored at various depths, the plates were processed below ground. The observed frequencies of μ -mesons stopped in the plates are consistent with those expected from measurements on the energy-

spectrum of cosmic rays. Several π -mesons stopping in the plates were observed, and reasons for believing these to have been locally produced in the matter near the plates are discussed.

Forty-two nuclear disintegrations have been observed, the frequency at a depth of 60 m. water equivalent being of the order of 5×10^{-3} stars/cm³/day. Approximately one third of these are attributed to the electromagnetic interaction of μ -mesons on their passage through nuclei. The remainder are attributed in part to neutrons from this first group of stars, and in part to the photons of the soft component underground.

Four examples of stars accompanied by showers of particles at minimum ionization have been observed, and are discussed.

Heat Changes Accompanying Magnetization in Low and Moderate Fields: the Effects of Strain, and a Theoretical Interpretation, by L. F. BATES and J. H. DAVIS.

ABSTRACT. The technique of Bates and his co-workers was used to measure the adiabatic temperature changes occurring during magnetization of nickel under conditions of progressively increased internal strain as a pure specimen was taken from the fully annealed state to the state of maximum strain, in absence of deforming force. A new technique enabled accurate thermal curves during virgin magnetization, and in cycles up to 400 oersteds, to be obtained; the induction effect previously reported was investigated and its origin determined. A new theoretical treatment is described in which account is taken of the energy changes associated with magnetostriction and which enables the thermal curves to be calculated from available data on magnetization and magnetostriction, and a general physical interpretation of the curves is given. Agreement is good for strained specimens and an explanation of the less satisfactory agreement for annealed specimens is offered.

The Luminescence of Diamonds, by C. BULL and G. F. J. GANTON.

ABSTRACT. Studies of the luminescence characteristics of different types of industrial diamonds (80-100 mesh dust) have been made. Blue fluorescent diamonds (3,650 Å. exciting radiation) exhibit thermoluminescence characteristics of two groups of metastable electronic states, with mean activation energies of 0.5 and 0.7 ev. respectively. Transitions of a forbidden nature from these states to the normal states of emission centres give rise to a temperature-independent, green-yellow phosphorescence, whose longest-lived component has a decay constant of 2×10^{-6} sec⁻¹. Most yellow luminescent and non-luminescent diamonds (3,650 Å. exciting radiation) exhibit a blue luminescence when excited by energetic particles or by x-rays, but no thermoluminescence is observed after any type of excitation. The luminescence of diamonds appears to be characteristic of the matrix crystal, and a tentative model for the emission centres is given.

THE PROCEEDINGS OF THE PHYSICAL SOCIETY

Section B

VOL. 63, PART 12

1 December 1950

No. 372 B

Quelques recherches sur les raies faibles dans les spectres optiques

By P. JACQUINOT

University of Paris

Fifth Holweck Discourse, delivered 17th May 1950

ABSTRACT. En plus de la faiblesse absolue des raies, c'est surtout leur faiblesse relative qui interdit leur observation lorsqu'elles sont trop voisines d'autres raies. Elles sont alors masquées par la lumière provenant de l'étalement des raies voisines, qui est dû essentiellement à la diffraction dans le spectrographe.

Il est souvent possible d'observer les raies faibles en modifiant dans la source elle-même les rapports d'intensité entre les différentes raies, grâce au phénomène d'autoabsorption, dont quelques exemples d'utilisation sont donnés.

On peut aussi modifier profondément la figure de diffraction classique de façon à en atténuer les 'pieds'. Cette 'apodisation' est obtenue, soit par des diaphragmes de formes diverses, soit par des écrans absorbants dégradés suivant des lois convenables. Quelques indications sont données sur le calcul de ces lois, sur la réalisation des écrans, et sur les résultats obtenus, consistant en des améliorations de contraste au voisinage des raies, dans un rapport atteignant facilement 10 000.

Le plus intéressant des résultats spectroscopiques obtenus jusqu'à présent est la mise en évidence et l'étude des raies d'intercombinaison de l'hélium. La plus facile à observer de ces raies se présente comme un satellite de la raie jaune, distant de 1 Å., et 10 000 fois plus faible. Il a même été possible d'observer l'effet Zeeman de cette raie; cet effet Zeeman est d'ailleurs différent de celui que l'on doit attendre, et cette contradiction pose un problème de rayonnement qui n'a pas reçu de solution.

I SHOULD like to say first how happy I am to be the fifth Holweck Prizewinner; it is for me a very great honour to have been recognized by the distinguished scientists who form the Physical Society Council.

Knowing the important part played by this prize in the friendly relations between scientists of our two countries, I am very proud to be the object of one of these expressions of friendship, the idea of which we owe to the eminent Foreign Secretary of the Physical Society. Professor Andrade's initiative will keep green for many years the memory of the life and death of Fernand Holweck, who was one of the most original and creative of our French physicists.

I knew Holweck a little. I admired his intuition, the acuteness of his way of thinking, and his experimental skill. I know too that his judgment of those of his colleagues whose work lacked character was sometimes a bit severe. I appreciate all the more the honour that is done to me today, because his name is attached to the Prize.

I am also very proud to see my name appear on the Holweck Prize list, which opened so brilliantly with those of Professor Andrade, of my old friend Charles Sadron, and of Professor Bates and Professor Rocard. My personal merits are rather modest, when compared with those of the former names, but it is impossible to suggest in public that there might be any doubt about the wisdom of the decisions of the Physical Society Council!

I have preferred to say a few words in English. I am sure now you will have no regret if I change language. You will surely understand me far better if I avoid going on torturing the language of Newton and Faraday.

Le sujet de ma conférence est une brève revue de ce que l'on doit faire, et de ce que l'on peut trouver lorsqu'on veut, par une curiosité peut-être excessive, chercher des raies spectrales situées à quelques dixièmes d'Angstrom d'autres raies quelques milliers de fois plus intenses qu'elles.

Les interféromètres se prêtent mal à de telles recherches. Cependant si les différences de longueur d'ondes deviennent très faibles ils sont les seuls appareils utilisables. Il y a en Angleterre d'éminents spécialistes de l'interférométrie qui savent combien on est alors limité par les questions de contraste. Kuhn a déjà obtenu dans ce sens des résultats intéressants par l'emploi de deux appareils Fabry-Perot en série, et Dufour (1949) a étudié le problème dans sa thèse récente. Mais je ne considérerai aujourd'hui que de l'emploi des spectrographes à prismes ou à réseau, c'est-à-dire ceux dans lesquels on forme l'image monochromatique d'une fente.

§ 1. LIMITE DE RÉOLUTION POUR DEUX RAIES D'INTENSITÉS TRÈS DIFFÉRENTES

Si une raie faible, ou un groupe de raies faibles, était isolée dans le spectre, à étudier il y aurait bien des moyens de la détecter malgré sa faiblesse, emploi de sources plus intenses, en particulier de sources longues, poses de grande durée dans le cas de la détection photographique, systèmes à grande constante de temps dans le cas de la détection photoélectrique. De toute façon, plus la raie est faible, plus il faut de moyens puissants, et de temps, pour l'étudier, mais si cette faiblesse absolue des raies exige beaucoup de travail, elle ne constitue pas un obstacle insurmontable.

Malheureusement les raies ne sont jamais seules, et les raies 'faibles' encore moins que les autres. Elles peuvent être mêlées à toutes sortes de lumières parasites, en particulier à des spectres d'impuretés, et tous les spectroscopistes connaissent bien les difficiles problèmes de purification que cela pose. Mais la recherche des raies faibles devient vraiment très difficile quand il se trouve dans leur voisinage des raies beaucoup plus intenses qu'elles, et qui appartiennent normalement au même spectre. Si la différence de longueur d'onde $\Delta\lambda$ des deux raies est trop faible, ou leur rapport d'intensité K trop grand, il devient impossible de détecter la raie faible, et si K est par exemple de l'ordre de 1000, $\Delta\lambda$ peut être 10 ou 100 fois plus grand que la limite de résolution conventionnelle $\delta\lambda$ définie pour deux raies d'égale intensité. Cela est dû au fait que la lumière s'étale autour de la position géométrique d'une raie même parfaitement monochromatique, beaucoup plus qu'on ne le croit d'habitude. Si l'on trace la courbe de répartition de l'éclairement dans l'image monochromatique d'une fente, cette courbe se compose d'une partie centrale ou 'corps', et d' 'ailes' ou 'pieds' beaucoup plus faibles, dont seule une représentation en ordonnées.

logarithmiques peut donner une idée correcte. On réserve le plus souvent le nom d'ailes aux parties de l'image relativement intenses et proches du corps de la raie, et le nom de pieds aux parties de faible intensité relative—par exemple inférieure à 5×10^{-3} du maximum et qui peuvent s'étendre très loin. Les aberrations géométriques provoquent généralement un élargissement du corps de l'image, et l'apparition d'ailes, souvent dissymétriques. Mais les pieds sont dûs le plus souvent à la plus essentielle des limitations de la définition, qui est la diffraction. J'ai souvenir qu'on n'enseigna que la figure de diffraction d'une fente donnée par un objectif rectangulaire de largeur a à la distance f était composée d'une frange centrale, de largeur $\lambda f/a$, et de franges latérales (figure 1)* de largeur deux fois plus petite, et d' 'intensité négligeable'. Mais

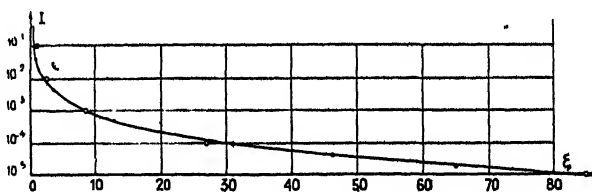


Figure 2. Loi de décroissance de l'éclairement de franges en fonction de leur numéro; ordonnées logarithmiques.

la dixième frange a encore une intensité relative de 10^{-3} (figure 2), ce qui est loin d'être négligeable quand les raies à séparer ont un contraste de 10^4 , et l'intensité des franges décroît seulement comme l'inverse du carré de leur numéro. Un calcul très simple montre alors que le pouvoir de résolution pour deux raies de contraste K serait ainsi, pour K grand, environ $\sqrt{K/\pi}$ fois plus faible que pour deux raies d'égale intensité, en supposant que la raie faible est résolue si son intensité est au moins égale à celle du pied de la raie forte à sa place.

Ces résultats simples sont vrais seulement si c'est la diffraction qui régit entièrement la formation des images: tels est le cas si, préférant la résolution à la luminosité, on adopte une distance focale f suffisamment grande pour que le 'grain' de la plaque, les largeurs de fente utilisées ou l'étendue des taches d'aberration soient au plus égales à la largeur $\lambda f/a$ de la tache centrale de diffraction. On dit alors que l'ouverture a/f est égale à l'ouverture résolvente, qui est de l'ordre de $1/40$ en général. Il en est bien ainsi dans les grands spectrographes à haut pouvoir de résolution utilisés pour étudier la structure des spectres d'émission atomiques et moléculaires: on dispose alors de tout le pouvoir de résolution intrinsèque du système dispersif.

Mais de tels appareils sont peu lumineux et on préfère souvent employer une ouverture m fois plus grande que l'ouverture résolvente: dans la cas des appareils photographiques on gagne alors en luminosité comme m^2 , et l'appareil possède alors un pouvoir de résolution réel m fois plus petit que le pouvoir de résolution intrinsèque.† Dans la cas de deux raies de contraste K , la relation est alors moins simple à établir entre le pouvoir de résolution R_k et le contraste K , mais une étude plus détaillée montre que R_k est $\sqrt{K/\pi}\sqrt{m}$ fois plus faible que le pouvoir de résolution R réellement obtenu avec deux raies d'égale intensité. On peut donc

* All photographs are reproduced on Plates I-IV, facing page 976.

† Le même résultat peut être obtenu dans le cas des appareils photoélectriques en agissant sur les largeurs de fente: le pouvoir de résolution obtenu est alors inversement proportionnel à la luminosité (Jacquinot et Dufour 1948).

dire que dans ce cas la diffraction joue relativement moins que dans le cas d'un spectrographe peu ouvert, mais elle joue cependant encore beaucoup trop.

Il semblerait donc que pour augmenter R_k on soit, de toute façon, obligé d'augmenter R . Mais si l'appareil travaille déjà avec son pouvoir de résolution intrinsèque R_0 , il n'y a pas d'autre moyen que d'augmenter les dimensions de l'appareil dispersif ou sa dispersion angulaire, ce qui est souvent impossible. Si, au contraire, l'appareil est assez ouvert pour que son pouvoir de résolution réel R soit inférieur à sa valeur intrinsèque R_0 , on peut augmenter R en diminuant l'ouverture, c'est à dire en choisissant une longueur focale m fois plus grande : on gagne alors un facteur \sqrt{m} sur R_k , mais en revanche on perd un facteur m^2 sur la luminosité, ce qui est très désavantageux. Un autre procédé consisterait à faire comme dans les monochromateurs doubles, où l'on a surtout en vue l'élimination de la lumière *diffusee*, beaucoup moins gênante en général que les 'pieds' de diffraction : après la traversée d'un premier appareil, la raie faible est isolée au moyen d'une fente et renvoyée dans un deuxième appareil, qui, dans ce cas, peut être le même grâce à un système optique convenable. Avec une telle disposition l'intensité relative des 'pieds' de la raie forte est élevée au carré, mais on ne peut ainsi déceler qu'une seule raie faible dont la place doit être connue à l'avance, et, en outre, la double traversée de l'appareil fait perdre beaucoup de lumière.

C'est à deux autres procédés que j'ai eu recours à Bellevue pour les recherches que j'y ai entreprises avec la concours de quelques collaborateurs dont les principaux sont J. Brochard, et M^{lle} Dossier. Le premier de ces procédés, qui n'est pas toujours applicable, consiste à modifier, dans la source lumineuse elle-même les rapports d'intensités des raies. Le second, plus général, consiste à modifier la figure de diffraction de façon à en atténuer les pieds : c'est ce que nous avons appelé 'l'apodisation'.

§2. MODIFICATION DES RAPPORTS D'INTENSITÉ PAR AUTOABSORPTION

On peut, dans une source lumineuse modifier soit les rapports d'intensité vrais des raies, soit les rapports de leurs intensités apparentes. De nombreux exemples de modification d'intensités vraies sont connus, et c'est ainsi que de nombreuses raies interdites ont pu être étudiées en favorisant l'accumulation d'atomes sur des niveaux de départ plus ou moins métastables. Et je ne veux pas parler ici des raies forcées, où l'on agit directement sur les probabilités de transition de raies qui doivent normalement ne pas exister. Mais il arrive souvent que l'on ne peut aucunement agir sur les populations relatives des différents niveaux, ni sur les probabilités de transition, mais où l'on peut agir sur les intensités relatives *apparentes* des raies émises par une source donnée. Cette action est possible grâce au phénomène d'autoabsorption des raies, souvent considéré comme gênant, sinon négligeable, et dont, le plus souvent, on omet prudemment de parler. Toute raie émise par un point d'une source est, en effet, partiellement absorbée dans la source même sur la parcours de la lumière et l'intensité apparente résulte de l'intégration combinée de l'émission et de l'absorption. Or les coefficients d'absorption sont dans quelques cas proportionnels aux coefficients d'émission (loi de Kirchhoff), et presque toujours, sans leur être aussi simplement liés, croissent en même temps qu'eux. Il en résulte que les raies intenses sont relativement beaucoup plus autoabsorbées que les raies faibles, ce qui provoque une diminution du contraste entre les intensités apparentes des raies (Jacquinot 1941). Cet effet est d'autant plus sensible que les coefficients d'absorption sont plus

grands, et que la source a une épaisseur plus grande dans le sens de la visée; cela conduit à employer des sources assez fortement excitées, et, surtout ayant la forme de tubes très longs. Dans un tube long, tout se passe donc comme si les raies intenses, donc fortement absorbées, n'étaient émises que par une tranche très mince à la sortie du tube, alors que les raies faibles non absorbées sont émises par le tube entier et ont une intensité proportionnelle à sa longueur. Le calcul est très simple à faire pour des raies supposées monochromatiques (ou de 'profil rectangulaire'). Si L est la longueur du tube, I_0L l'intensité qu'aurait une raie si l'absorption était nulle, et a le coefficient d'absorption, l'intensité apparente est $I = (I_0/a)(1 - e^{-aL})$, qui se réduit à I_0L si a est très faible et à I_0/a si L est grand. Le contraste entre une raie faible et une raie forte se trouve donc divisé par un facteur $G = aL$, qui est le gain dû à l'autoabsorption.

Le même résultat est encore valable en chaque région d'une raie non monochromatique dont la largeur serait nettement plus grande que la limite de résolution du spectrographe. Mais pour une raie réelle dont la largeur est plus faible que cette limite, le calcul doit comporter une intégration sur toute la raie, ce qui nécessite une connaissance exacte de la *forme réelle* de la raie, c'est à dire de la loi suivant laquelle varient les coefficients d'émission et d'absorption en fonction de la longueur d'onde. Or on ne peut se contenter de faire des hypothèses sur cette forme de raie, par exemple d'admettre que la répartition d'intensité est due seulement à l'effet Doppler, car, si l'autoabsorption est forte, même des intensités très faibles à grande distance du centre de la raie peuvent fausser les résultats. De toute façon bien que le calcul soit souvent impossible, le gain G de contraste dû à l'autoabsorption, est moins grand que celui que l'on pourrait calculer pour une raie monochromatique.

Les gains que l'on obtient peuvent être considérables. C'est ainsi qu'avec des tubes de 2 mètres de longueur, on obtient pour certaines raies de l'hydrogène ou de l'hélium des diminutions de contraste dans un rapport atteignant 100. Si l'on observe dans un tel tube la raie rouge $H\alpha$ de l'hydrogène, la raie du deuterium apparaît seulement 50 fois plus faible, bien que la teneur en deuterium ne soit que de 1/5 000.

Malheureusement cette méthode d'autoabsorption n'est pas très générale: toutes les raies ne sont pas aussi facilement absorbables, et l'emploi de sources très longues est très difficile pour produire les spectres des éléments non gazeux aux températures moyennes. En outre il est souvent très difficile de connaître dans quel rapport ont été modifiées les différentes intensités relatives, et s'il est ainsi possible de mettre en évidence des raies nouvelles, la mesure de leurs intensités relatives vraies peut être très compliquée, et quelquefois impossible.

C'est en partie grâce à cette méthode d'autoabsorption que nous avons pu à Bellevue poursuivre de nombreuses études sur des raies interdites de l'hélium et de l'hydrogène. L'étude de l'autoabsorption sera, en outre, poursuivie pour elle-même, car elle permet d'obtenir de nombreux renseignements sur les populations des niveaux et les probabilités de transition: les mesures existantes actuellement à ce sujet sont en effet souvent discordantes et il arrive de voir deux auteurs différents donner d'un même coefficient d'absorption deux valeurs dans le rapport de 1 à 50.

Si l'on ne peut modifier les intensités relatives apparentes des raies, ou si cette modification est insuffisante, il faut s'attaquer au phénomène de diffraction lui-même. Là encore nous distinguerons deux méthodes, dont l'une utilise une fente réduite à un point et l'autre une fente ordinaire.

§ 3. APODISATION PAR ÉCRANS GÉOMETRIQUES

Les astronomes, qui travaillent souvent avec des sources ponctuelles, connaissent bien la méthode qui permet de voir, par exemple, le compagnon de Sirius. On remplace le contour circulaire de l'objectif par un contour polygonal : les anneaux de diffraction sont alors remplacés par une figure étoilée, et l'éclairement décroît, entre les branches, plus rapidement que dans le cas des anneaux. Cette méthode peut être transposée à la spectroscopie, en utilisant une fente réduite à un point et un contour d'objectif polygonal tourné de telle sorte que l'axe de dispersion soit bissectrice entre deux branches de la figure de diffraction étoilée. C'est cette méthode que nous avons, avec A. Couder, proposée en 1939 (Couder et Jacquinot 1939), en employant un contour carré dont les cotés font un angle de 45° avec l'arête du prisme. La figure de diffraction est celle de la figure 3 : l'éclairement y décroît, suivant les branches de la croix comme $1/x^2$ mais, suivant les diagonales, comme $1/x^4$. L'inconvénient du procédé est qu'on ne peut utiliser qu'une fente, sinon 'ponctuelle', du moins très courte, dont la longueur ne dépasse pas de 5 à 50 fois sa largeur suivant la distance de la raie faible cherchée. Cette limitation de la longueur de la fente est très gênante, car une véritable raie se voit mieux qu'un point, et la mesure des distances est plus difficile. En outre la longueur de la fente, qui n'intervient pas dans la luminosité 'photographique', intervient directement dans le cas de l'exploration photoélectrique directe du spectre.

Cependant l'emploi de cette méthode permet d'obtenir très vite des résultats nouveaux : la figure 4 montre l'exemple de raies très faibles au voisinage de la raie bleue 4471 Å. de l'hélium, et qui ont été littéralement *devoilées*, car sans le diaphragme les branches de la croix auraient été rabattues sur elles.

On peut encore accroître l'efficacité de la méthode et mieux dégager dans une direction les abords du centre de la figure de diffraction en utilisant des contours de formes mieux étudiées. Le calcul complet de la figure de diffraction d'un diaphragme de forme quelconque est assez laborieux, mais il se simplifie si l'on considère des types tels que ceux de la figure 5 et si l'on ne s'intéresse qu'à l'éclaire-

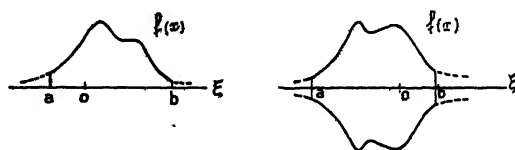


Figure 5. Contours de diaphragmes.

ment sur l'axe de symétrie $O\xi$, parallèle à Ox , de la figure de diffraction : l'éclairement s'obtient alors par une intégrale simple de Fourier appliquée à la fonction $y = f(x)$ entre a et b . J'indiquerai plus loin comment on peut déterminer $f(x)$ pour que la figure de diffraction présente les qualités voulues : je me contenterai seulement ici de montrer quelques figures de diffraction dans lesquelles la lumière diffractée a été déplacée de la région intéressante beaucoup mieux qu'avec un simple contour carré ou polygonal (figures 6, 7, 8).

Evidemment ces diaphragmes 'géométriques' présentent toujours, pour la spectroscopie, l'inconvénient de nécessiter une fente très courte, mais ils peuvent avoir d'autres applications. La figure 9 montre, en particulier, leur application à l'étude de la lumière diffractée par un défaut d'homogénéité du verre ('fil' perpendiculaire à Ox), étude faite par M. Françon, de l'Institut d'Optique.

§ 4. APODISATION PAR ÉCRANS ABSORBANTS

Pour modifier la figure de diffraction tout en employant une fente comme source il faut agir, non plus sur le contour de l'objectif, ou plus généralement de la pupille de sortie du spectroscope, mais sur la répartition des amplitudes sur cette pupille. Ces répartitions sont obtenues en disposant sur la pupille un écran absorbant dont la transmission varie d'un point à un autre. En particulier si l'amplitude lumineuse, au lieu de s'annuler brusquement à la limite de la pupille, s'annule progressivement, toute périodicité disparaît dans la figure de diffraction. Mais ce qui est intéressant n'est pas de supprimer la périodicité, c'est d'obtenir des éclaircissements relatifs aussi faibles que possible tout près du centre de l'image : tel est le problème de l'apodisation'.

Je ne développerai pas le traitement mathématique de la question, exposé en détail par ailleurs (Dossier, Boughon et Jacquinot 1950), mais je veux seulement indiquer pourquoi il n'est pas simple. Le cas le plus simple, et aussi le plus utile pour le spectroscopiste, est celui d'une pupille de contour rectangulaire dont un axe, Oy, est parallèle à la fente, l'amplitude lumineuse $f(x)$ étant constante le long de toute parallèle à Oy ; l'éclairement $I(\xi)$ dans la figure de diffraction alors donnée par le carré de la transformée de Fourier de $f(x)$, fonction nulle en dehors de la pupille. Si, réciproquement, on se donne arbitrairement $I(\xi)$ on trouvera une 'fonction pupillaire' $f(x)$, qui, généralement, ne sera pas nulle en dehors de la pupille. Le problème réciproque n'a donc, en général, aucune solution. On doit se contenter d'imposer à la 'fonction d'image' $I(\xi)$ certaines conditions de concentration ou de décroissance de l'énergie dites 'critères d'apodisation' : les équations intégrales auxquelles on aboutit ainsi ne peuvent être résolues que par des méthodes d'approximation. La méthode adoptée consiste à exprimer $f(x)$ suivant un polynôme de Fourier à deux ou trois termes et à en déterminer les coefficients de façon à satisfaire au critère d'apodisation.

Suivant les critères choisis et la largeur d'image dans laquelle on cherche à concentrer l'énergie on aboutit à des fonctions différentes donnant des images plus ou moins apodisées, dont la partie centrale ou 'corps' est plus ou moins élargie. En général, plus on veut une apodisation prononcée, plus il faut consentir à élargir le corps de l'image et à perdre de la lumière. Il n'existe pas *une* solution meilleure que les autres, mais des solutions différentes suivant le problème posé, recherche d'un satellite faible et lointain, ou très faible et proche. A titre d'exemple, les caractéristiques d'une solution moyenne sont données dans la figure 10 et dans la tableau, comparées à celle de la pupille de dimensions identiques et de transmission uniforme : cette comparaison a été établie en tenant compte de la largeur finie de la fente, choisie dans chaque cas de façon à assurer le meilleur compromis entre la luminosité et le pouvoir de résolution.

Caractéristiques d'un écran apodiseur.

Loi de transmission :	$f(x) = 0,4935 + 0,4865 \cos \pi x + 0,0200 \cos 2\pi x$ (écran C_2)					
Perte de pouvoir résolvant :	30%					
Perte de luminosité théorique :	50%					
Coefficient d'apodisation $K(\xi)$	ξ	2,5	4,5	6,5	9,5	$+\infty$
	$K(\xi)$	270	5,550	830	482	336

ξ est le rapport de l'abscisse à la limite de résolution conventionnelle.

Le coefficient d'apodisation $K(\xi)$ est le rapport, en un point ξ donné, de l'éclairement relatif sans écran à l'éclairement relatif avec écran.

On voit que, bien que l'apodisation soit considérable, la perte de pouvoir résolvant au sens conventionnel est seulement de 30 % et la perte de lumière de 50 %.

La figure 11 montre les figures de diffraction comparées d'une ouverture rectangulaire avec et sans écran apodisant. Les cotés du rectangle sont dans le rapport de 1 à 5, les lignes d'égale opacité de l'écran apodiseur étant parallèles aux petits cotés (Oy). Les franges de diffraction suivant O α sont supprimées, alors que les autres, 5 fois plus larges sont inchangées. L'élargissement du corps de l'image, visible sur les franges verticales, est ici de l'ordre de 1,7.

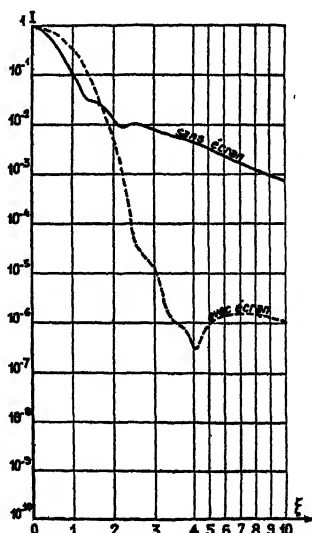


Figure 10. Loi de décroissance de l'éclairement avec l'écran apodiseur C_2 (voir tableau), et une largeur de fente source égale à la largeur d'une frange de diffraction.

Des coefficients d'apodisation de même ordre de grandeur sont obtenus dans le cas des appareils d'ouverture plus grande que l'ouverture résolvante: on peut donc obtenir une plus forte amélioration de contraste, et au prix d'une perte de lumière beaucoup plus faible qu'en diminuant l'ouverture jusqu'à sa valeur résolvante. Dans ce cas, d'ailleurs, l'apodisation n'est accompagnée d'aucune diminution du pouvoir résolvant, puisque celui ci est fixé uniquement par la limite de résolution de la plaque photographique.

Ces écrans apodisants peuvent être réalisés par dépôts de métaux évaporés dans le vide. La variation d'épaisseur suivant la loi désirée s'obtient par déplacement de l'écran, pendant le dépôt, derrière un cache de forme convenablement calculée.

Ce déplacement doit s'effectuer d'un mouvement alternatif uniforme, ce qui entraîne certaines complications mécaniques: un appareil (figures 12 et 13) a été construit spécialement pour la préparation de ces écrans. La préparation d'écrans ayant la symétrie de révolution serait beaucoup plus simple, mais les écrans de révolution sont moins bien adaptés au cas du spectrographe.

§ 5. DEFECTS DE L'APPAREIL DISPERSIF

Les différentes méthodes d'apodisation n'agissent que sur le phénomène de diffraction et elles perdent toute efficacité si le système optique possède des défauts

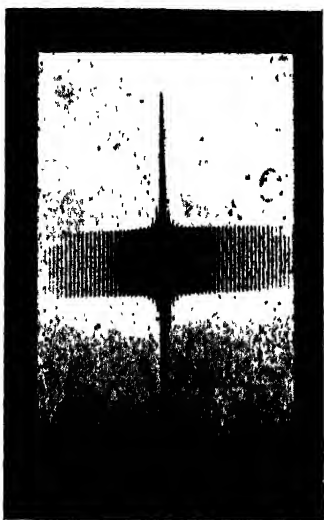


Figure 1. Photographie des franges de diffraction d'une fente.



Figure 4. Raies satellites de la raie He 4471, qui apparaissent grâce à l'utilisation d'un diaphragme carré tourné de 45° par rapport à la direction de dispersion.

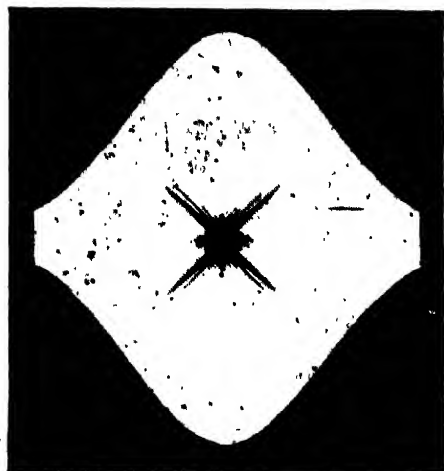


Figure 7. Même figure que 6 avec $f(x) = \exp(-2x^2)$.

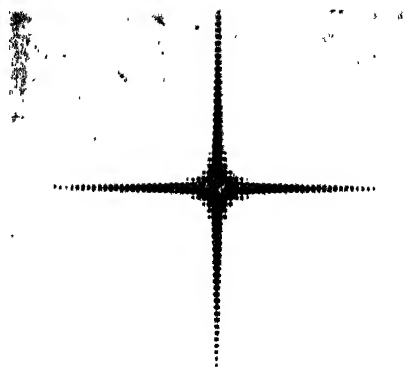


Figure 3. Figure de diffraction d'un point par une ouverture carrée.

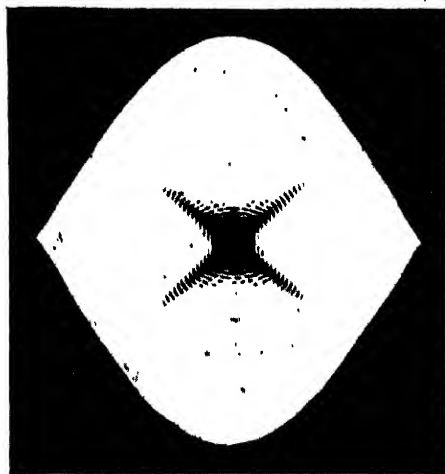


Figure 6. Figure de diffraction d'un point donnée par un diaphragme de contour $f(x) = \cos \frac{1}{2}\pi x$. Le contour de diaphragme a été dessiné de façon à encadrer la figure de diffraction.

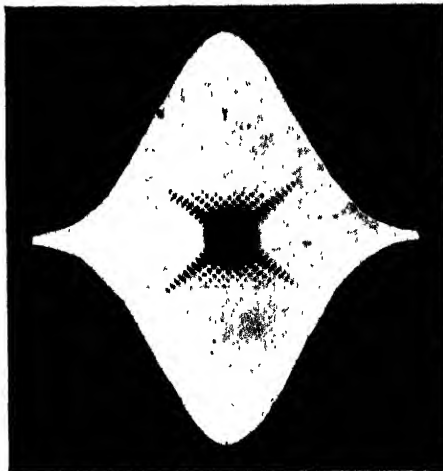


Figure 8. Même figure que 6 avec $f(x) = 0,4935 + 0,4865 \cos \pi x + 0,0200 \cos 2\pi x$.



Figure 9.

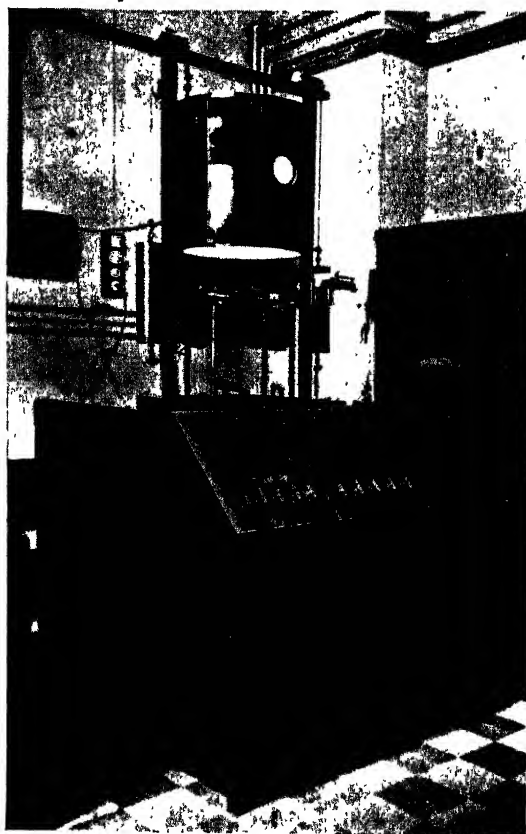


Figure 12. Appareil d'évaporation pour la préparation des écrans,

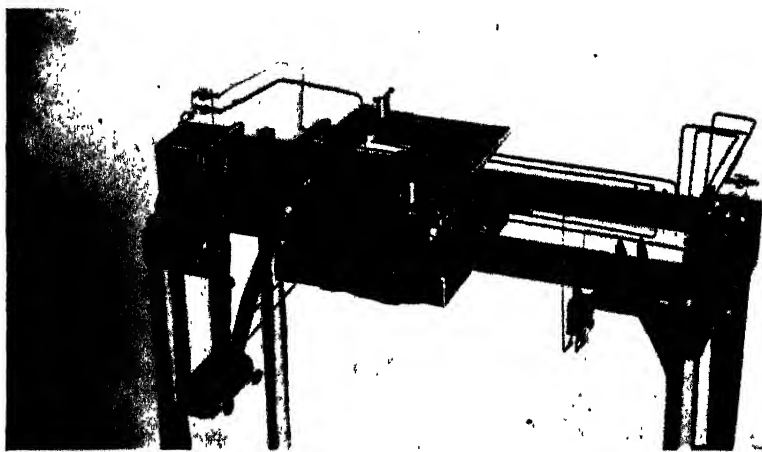


Figure 13. Détail du dispositif permettant d'obtenir un mouvement rectiligne alternatif uniforme.



Figure 14. Lumière parasite produite par les irrégularités d'un réseau de Rowland; on distingue, en plus, les ghosts habituels dus aux erreurs périodiques.

PLATE III.



Figure 11. Image d'un point à travers un écran rectangulaire uniforme (à gauche) et à travers un même écran apodisant dans une direction (à droite).

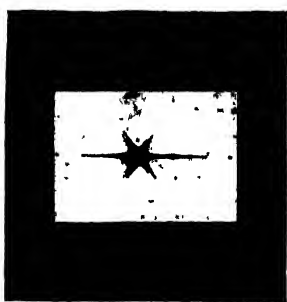


Figure 15. Pieds 'spectraux' de la raie 4471 de l'hélium, qu'on peut supprimer en abaissant la pression.

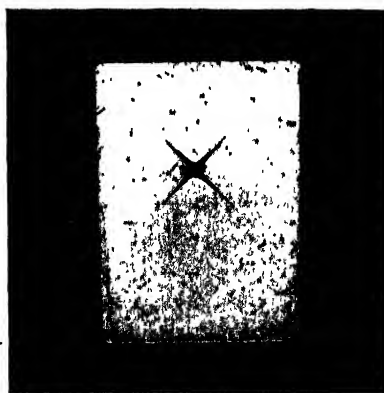


Figure 16. La raie jaune 5875 et la raie d'intercombinaison qu'on distingue juste entre les bras de la croix, à droite.

produisant des pieds d'intensité comparable à ceux de la figure normale de diffraction : ces défauts passent généralement inaperçus car on ne les sépare pas de la diffraction normale. Un bon moyen de les mettre en évidence est de faire une photographie très surexposée avec une fente de très faible hauteur et l'un des diaphragmes géométriques décrits précédemment. Ce contrôle a été appliqué à un réseau concave de Rowland de 21 pieds et le résultat en est montré sur la figure 14 obtenue avec le raie verte du mercure dans le premier ordre avec un diaphragme en forme de losange. La lumière parasite, presque aussi gênante que la lumière diffractée qui est ainsi mise en évidence, est due aux irrégularités des traits du réseau, et il n'y a aucun moyen de la diminuer ; l'effet est encore plus important dans les ordres plus élevés. Ce réseau est donc impropre aux recherches sur les satellites faibles.

Je n'ai pas fait d'essais sur d'autres réseaux, car, heureusement, j'ai eu à ma disposition un remarquable instrument à prisme, construit en 1935, sur la demande de M. Cotton, par A. Couder. C'est un gros prisme à liquide, associé, en montage autocollimateur, à une lentille de 9 mètres de distance focale. La section droite du faisceau sortant du prisme à 20 cm. de diamètre et l'ensemble a, dans le bleu, un pouvoir résolvant comparable à celui d'un grand réseau de Rowland dans le deuxième ordre, sans en avoir aucun des inconvénients : la figure 16 montre une raie très fortement surexposée (environ 10 000 fois), donnant une lumière parasite très faible en comparaison de la lumière diffractée. C'est grâce à cet instrument, probablement unique, que j'ai pu obtenir la plupart de mes résultats sur les raies faibles.

§ 6. EXEMPLE DES RAIES D'INTERCOMBINAISONS DE L'HÉLIUM

Tous ces procédés, qui permettent l'observation de raies faibles au voisinage de raies intenses peuvent être appliqués à de nombreux problèmes, tels que la recherche des raies d'isotopes rares, l'étude de la forme des raies loin de leur centre, là où l'intensité due à l'effet Doppler est négligeable devant l'intensité, même très faible, due aux autres causes d'élargissement ; on peut aussi, chemin faisant, avoir l'heureuse chance de rencontrer des raies nouvelles auxquelles on ne songeait pas.

Et, pour en étudier l'exemple le plus intéressant, je vais raconter l'histoire de la raie jaune de l'hélium.

J'ai commencé à étudier cette raie en 1939 : c'était alors pour rechercher un isotope de masse 5 dont on pouvait penser, à cette époque, qu'il était stable quoique très peu abondant. Le développement de toutes les techniques qui viennent d'être exposées a son origine dans cette recherche.

J'étudiai donc quelques raies de l'hélium avec un tube de 20 cm. de longueur et le grand prisme muni d'un diaphragme carré à 45°, qui était le seul procédé que je connaisse alors. En surexposant les photographies j'obtins tout d'abord des étoiles à 6 branches (figure 15) au lieu des étoiles à 4 branches que j'attendais : c'est que les raies présentaient elle mêmes des pieds dûs non pas à la diffraction, mais au phénomène même de l'émission. Ces pieds, très faibles, échappent aux observations habituelles où on peut les confondre avec la lumière diffractée car ils décroissent suivant la même loi. Il suffit de diminuer suffisamment la pression dans le tube pour faire disparaître ces pieds 'spectraux'. Les abords de la raie se trouvent alors suffisamment dégagés pour que l'observation de raies faibles y soit possible : les figures 16 et 4 relatives aux raies 5875 ($2^3\text{P}-3^3\text{D}$) et 4471 ($2^3\text{P}-4^3\text{D}$) montrent les raies qui apparaissent alors et dont l'intensité est de l'ordre du dixmillième de celle de la raie qu'elles accompagnent.

Parmi toutes les raies qui apparaissent ainsi (Jacquinet 1939), certaines sont des raies 'forcées' dont l'intensité relative augmente avec la densité de courant, et sont dues aux champs interioniques : il a été possible par la suite, avec des moyens plus puissants—tubes plus poussés et utilisation de l'autoabsorption—d'en étudier complètement l'effet Zeeman (Brochard et Jacquinet 1945), et cette étude a conduit à des résultats curieux sur la répartition des champs électriques interioniques qui régissent à l'intérieur d'une décharge placée dans un champ magnétique.

D'autres raies ont une intensité relative *vraie* indépendante des conditions d'excitation : tel est le cas de celle qui accompagne à 1,12 Å. de distance la raie jaune 5875 (2^3P-3^3D), de celle qui accompagne à 1,92 Å. de distance la raie rouge 6678 (2^1P-3^1D) et de leurs homologues supérieures dans ces deux séries. Malgré leur faible intensité relative vraie ces raies s'observent bien, même sans dispositif apodiseur grâce à l'autoabsorption, avec un tube de 2 mètres, ainsi que le montre l'enregistrement de la figure 17 obtenu directement avec un dispositif photo-électrique sensible placé dans le plan focal du grand spectrographe à prisme. Leur probabilité de transition relative a pu être mesurée exactement par M. Brochard (1949) en tenant compte de l'effet d'autoabsorption. On identifie facilement ces raies comme des intercombinaisons entre niveaux de singulets et niveaux de triplets conformément à la figure 18. On ne connaissait jusqu'alors que

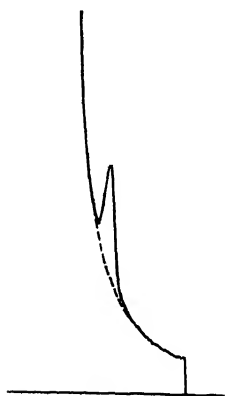


Figure 17. Enregistrement direct de la raie rouge d'intercombinaison de l'hélium.

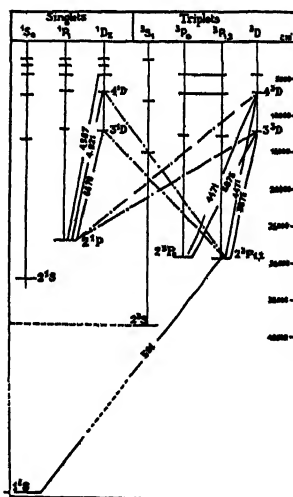


Figure 18. Schéma du spectre de l'hélium, montrant les nouvelles raies d'intercombinaison.

deux ou trois raies d'intercombinaison, situées dans l'extrême ultraviolet : encore ces raies étaient-elles contestées à cause de la démonstration, donnée par Heisenberg, de l'impossibilité de ces combinaisons si l'on ne tient pas compte du spin. Cependant ces raies ne sont pas complètement interdites, à cause des interactions spin-orbite et comme la transition a lieu entre deux niveaux de parités différentes, la règle des parités de Laporte permet leur existence en qualité de raies dipolaires électriques : par contre cette même règle les interdit rigoureusement pour le rayonnement quadrupolaire électrique.

L'étude de l'effet Zeeman, à condition de faire à la fois des observations transversales et *longitudinales*, doit permettre de savoir à quel type de rayonnement

elles correspondent. Mais, pour ces raies qui sont déjà difficiles à observer, l'étude de l'effet Zeeman est à limite des possibilités; c'est pourquoi il n'a pu être observé que sur la raie jaune (Brochard et Jacquinet 1946). Les clichés obtenus sont cependant suffisamment nets pour que l'on puisse affirmer qu'il s'agit d'une raie quadrupolaire, contrairement à ce que l'on devait attendre.

Cette raie voile donc la règle des parités, qui, par sa nature même doit être rigoureuse et n'est violée, je crois, par aucune autre raie connue. La raison de cette anomalie n'est pas encore apparue, et aucune explication théorique ne semble possible actuellement. Il existe des raies homologues, dans des atomes plus lourds, telle la raie 5770 ($2^1P-3^3D_2$) du mercure, dont l'intensité relative est très grande; mais cette raie, dont l'effet Zeeman est bien connu, est dipolaire et satisfait à la règle des parités. Nos recherches tendent actuellement à rendre possible l'étude de l'effet Zeeman des autres raies d'intercombinaison du type P-D de l'hélium, et des mêmes raies dans les éléments suivants à deux électrons optiques. Ces études présenteront sans doute de grandes difficultés, ne serait ce que parce que les raies d'intercombinaison homologues de celles de l'hélium ne sont pas encore connues dans les éléments qui suivent l'hélium, tel le beryllium. C'est afin de les rendre possibles que nous avons poursuivi l'étude de l'apodisation par écrans absorbants; ces écrans, à cause de nombreuses difficultés de réalisation n'ont pas encore été utilisés sur le grand spectrographe de Bellevue, mais ils le seront très prochainement. En outre le spectrographe est maintenant équipé d'un système à double faisceau, pour l'enregistrement direct des raies par voie photo-électrique, qui permet d'étudier même des raies extrêmement faibles.

BIBLIOGRAPHIE

- BROCHARD, J., 1949, *Thèse*, Paris.
 BROCHARD, J., et JACQUINET, P., 1945, *Ann. Phys., Paris*, 20, 509; 1946, *C. R. Acad. Sci., Paris*, 223, 507.
 COUDER, A., et JACQUINET, P., 1939, *C. R. Acad. Sci., Paris*, 208, 1639.
 DOSSIER, B., BOUGHON, P., et JACQUINET, P., 1950, *J. Recherches, C.N.R.S.*, No. 11.
 DUFOUR, C., 1949, *Thèse*, Paris.
 JACQUINET, P., 1939, *C. R. Acad. Sci., Paris*, 208, 1896; 1941, *Ibid.*, 212, 537.
 JACQUINET, P., et DUFOUR, C., 1948, *J. Recherches, C.N.R.S.*, No. 6.

Electron Microscopic Studies on Aqueous Sols

By M. R. A. RAO

Department of Chemical Technology, Imperial College, London

Communicated by G. I. Finch; MS. received 21st August 1950

ABSTRACT. Electron micrographs are shown for aqueous sols containing the following colloidal suspensions: graphite, stearic acid, vanadium pentoxide, and sulphur. The results obtained explain certain divergences noticed during the measurement of depolarization of light scattering in these colloidal solutions.

§ 1. INTRODUCTION

IT is generally assumed that colloidal graphite in aqueous sols consists of laminated anisotropic particles. Subrahmanya *et al.* (1944) have shown that colloidal stearic acid also consists of disc-shaped particles similar to colloidal graphite. In a recent study on the effect of an electric field on the degree of depolarization of light scattering in colloidal solutions, Subrahmanya and Rao (1949) have shown that the variation of the depolarization factor for the vertically and horizontally polarized light on the application of an electric field indicates a similarity between colloidal graphite and colloidal stearic acid. But with unpolarized light the two sols behave differently, showing thereby that the shape of the particles in the two cases may not be similar. The stearic acid in this case is similar to vanadium pentoxide sol, which consists of elongated particles. The twinkling of the colloidal particles and the schlieren experiments can only indicate the anisotropy of the particles but not the shape. An electron-microscopic study of these sols has therefore been undertaken with a view to determining the shape of the colloidal particles.

§ 2. EXPERIMENTAL

The preparation of colloidal graphite, stearic acid and vanadium pentoxide has been described in a previous paper (Subrahmanya and Rao 1949). Colloidal sulphur was prepared by the oxidation of aqueous hydrogen sulphide with sulphur dioxide. All these had to be suitably diluted before evaporating on the grid.

Of the several materials tried for the supporting film, 'Formvar' proved to be the best. The 'Formvar' membrane was removed on to the grid and dried. A drop of the colloidal solution was then placed on the membrane and evaporated slowly. Rapid evaporation had to be avoided to prevent the coagulation of the colloid. After drying, the particles were gold-shadowed at grazing incidence from a distance of 6 cm. from the gold source.

The electron microscope used in this investigation was constructed in Professor G. I. Finch's laboratory in the Imperial College of Science and Technology, London. It is a three-stage instrument worked at 80 to 100 kv. with a resolving power of about 200 Å.

§ 3. RESULTS

Figure 1* indicates a typical electron micrograph for colloidal graphite. This consists of two micrographs taken for different samples of the colloid. The sample

* For all Figures see Plate.

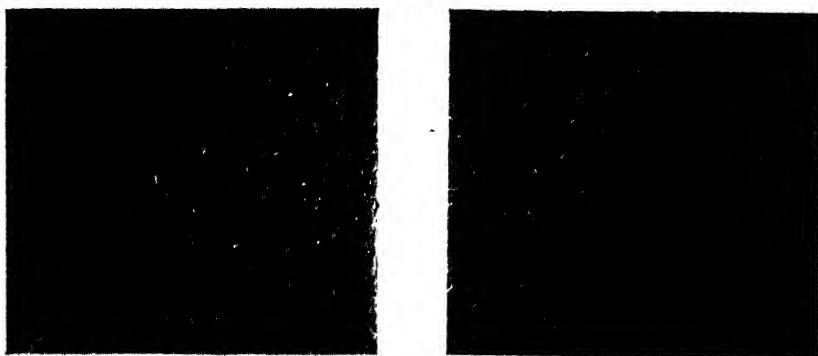


Figure 1. Colloidal graphite shadow cast with gold. $\times 5000$.

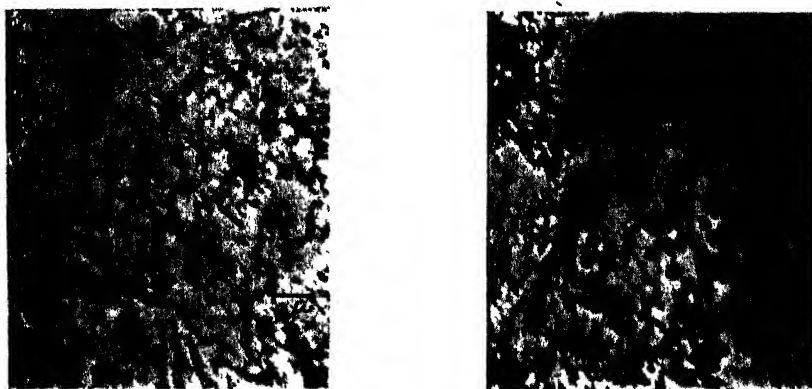


Figure 2. Colloidal stearic acid shadow cast with gold. $\times 5000$.

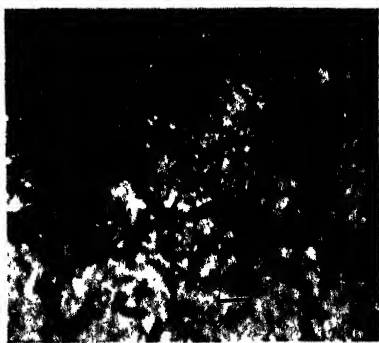


Figure 3. Colloidal vanadium pentoxide, $\times 5000$.

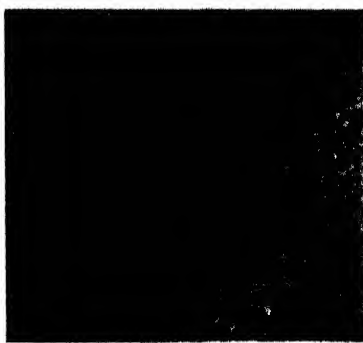


Figure 4. Colloidal sulphur. $\times 5000$.



consists of particles of varying sizes. An ultra-microscopic determination of the size of the particles gave a value of 0.138μ for the average diameter. The point of interest that has to be specially noted in this figure is the small length of the shadow cast by the particles. This indicates definitely that the colloidal graphite is laminated and that the thickness of the lamina is a small fraction of the average diameter.

Figure 2 is the electron micrograph for stearic acid. Unlike the graphite, the particles of stearic acid seem to have all kinds of shapes. The average diameter as determined by the ultra-microscope is 0.236μ , i.e. about twice that of colloidal graphite. The shadows cast by the particles in this case are much longer than in the previous case. This indicates that the thickness of the particles is much greater than that of the colloidal graphite. The electron microscopic study therefore clearly indicates that there is a vast difference in shape between colloidal graphite and colloidal stearic acid. In fact stearic acid seems to be a mixture of elongated and laminated particles, coming midway between graphite and vanadium pentoxide (Figure 3). The divergence in shape between graphite and stearic acid is responsible for the difference in behaviour of the two sols when the degree of depolarization is measured under the influence of the electric field (Subrahmanya and Rao 1949).

Figure 3 was obtained from freshly prepared vanadium pentoxide. It is interesting to note the presence of the two types of particles. The larger ones are about 1μ in length and 0.3μ in diameter, while the fine structure exhibits a thread-like appearance. On standing, the sol gives only the fine fibrous structure, as has been noticed by Huber and Zbinden (1949). This indicates that the large particles of the freshly prepared vanadium pentoxide undergo spontaneous disintegration on keeping.

A micrograph of colloidal sulphur is shown in Figure 4. This picture is similar to the electron micrograph of colloidal gold. The particles appear to be more or less spherical and very much smaller in diameter. Colloidal sulphur does not indicate any change in the degree of depolarization with orientation of the particles.

The coagulation of the particles during the evaporation of the sol on the 'Formvar' membrane may next be considered. The average diameter for the particle size obtained by the electron microscopic method is close to the value obtained with the ultra-microscope. This indicates that very little coagulation takes place when the sol is evaporated on the 'Formvar' membrane. The high dilution of the sol, the like charges on the colloids, and the presence of a stabilizing agent go a long way towards inhibiting coagulation.

ACKNOWLEDGMENTS

The author is deeply grateful to Professor G. I. Finch for valuable guidance and for providing laboratory facilities. His thanks are also due to Mr. M. Blümm for help in using the electron microscope.

REFERENCES

- HUBER, K., and ZBINDEN, H., 1949, *Z. anorg. Chem.*, **258**, 188.
SUBRAHMANYA, R. S., DOSS, K. S. G., and RAO, B. S., 1944, *Proc. Indian Acad. Sci.*, **29**, 1.
SUBRAHMANYA, R. S., and RAO, M. R. A., 1949, *Proc. Indian Acad. Sci. A*, **29**, 1.

Photoconductivity in Magnesium Antimonide Layers

By T. S. MOSS

Department of Colloid Science, Cambridge

MS. received 17th February 1950, and in amended form 1st May 1950

ABSTRACT. Layers of Mg_3Sb_2 evaporated *in vacuo* are found to exhibit semiconducting properties, although both constituents are normally metals.

Under suitable conditions the layers are photosensitive. Measurements have been carried out on the spectral distribution of sensitivity, and show bands with long wavelength edges at 1.5μ , 2.6μ and 3.5μ . The former is identified with intrinsic conductivity, and the two latter with impurity levels.

Activation energies found from resistance-temperature measurements are very similar to the optical activation energies corresponding to the above wavelengths.

§ 1. INTRODUCTION

UNTIL recently only two classes of semiconductor have been recognized, namely: (i) elements, (ii) compounds having as anion a member of group VI of the periodic table, i.e. oxides—tellurides.

A previous paper (Moss 1950) has shown that some compounds having a member of group V as the anion should have semiconducting properties, and that, as a result of the very large polarizabilities of these anions, such compounds should have interesting photoconductive properties. For example, the As^{---} ion has twice the polarizability of the Te^{--} ion (Pauling 1927a) with the same ionic radius (Pauling 1927b).

In order to produce photosensitive films of such compounds by vacuum evaporation methods, it is desirable that the vapour pressures of the component elements should be roughly equal, so that if some degree of dissociation of the compound occurs the components will evaporate at similar rates and will condense in nearly stoichiometric proportions. From this point of view the most suitable element to use in combination with arsenic is cadmium, and the properties of Cd_3As_2 layers are described by Moss (1950).

The present paper describes the extension of the work to cover antimonides. As regards vapour pressure, the best element to combine with antimony is magnesium (Landolt-Bornstein 1936). Also from the phase diagram of the Mg-Sb system a definite compound Mg_3Sb_2 seems to be formed (Hansen 1936). Accordingly the preparation and investigation of this compound was undertaken.

Soon after the work was started a paper by the Russian workers Boltaks and Zhuze (1948) appeared, giving measurements on the electrical properties of this compound, and verifying that it is indeed a semiconductor.

§ 2. PREPARATION OF BULK MATERIAL AND PHOTOSENSITIVE LAYERS

To form the compound, accurately stoichiometric proportions of 'Specpure' materials were baked together in a sealed-off evacuated quartz tube (pressure $\sim 10^{-6}$ mm. Hg on ionization gauge) at $1,050^\circ\text{C}$. for $2\frac{1}{2}$ hours.

All cells were made of Pyrex glass in the form of small Dewar flasks. Electrodes of either graphite or platinum about 5 mm. wide with 1 mm. separation were painted on the bottom of the inner part of the Dewar flask.

Early attempts to form layers by evaporation in an oven at temperatures below the softening point of Pyrex (i.e. $\sim 600^\circ\text{C}$.) always resulted in layers with metallic properties. Experiments were therefore carried out with evaporators built inside the cells. Small quartz crucibles of 1.5 mm. diameter and 3 mm. length were made to hold the Mg_3Sb_2 . For a typical layer 1 to 2 mg. of material were placed in a crucible, which fitted into a tungsten helix placed 1.5 to 2.0 cm. from the electrode surface.

A number of cells were made using various evaporation rates. Two of these had resistances of over 2×10^4 ohms at room temperature and tens of megohms at liquid air temperature, and showed some decrease in resistance on illumination whilst cool. However, on warming to room temperature the layers lost their semiconducting properties and became metallic in nature. In view of these results, subsequent cells were not in general subjected to such strong cooling.

The next cell produced was slightly photosensitive both at room temperature and at 195°K . Spectral sensitivity measurements carried out at both temperatures showed there was some sensitivity at wavelengths beyond the beginning of the Pyrex absorption region. Accordingly subsequent cells were made with 'bubble' windows of thin Pyrex, such as would give good transmission up to 4.5μ . Most of these later cells were sensitive, and the detailed measurements carried out on four of them are described.

§3. CELLS A AND B

For cell A the layer was formed in high vacuum: the cell was evacuated, the system (including ionization gauge) being pumped and outgassed until the gauge indicated 0.8×10^{-8} mm. Hg, and the cell then outgassed at 250°C . for an hour. After cooling, the material was evaporated (in about 3 seconds) to form a layer of $85\text{ k}\Omega$ resistance. The layer was opaque and had a metallic appearance.

For cell B the layer was evaporated at backing pressure, (approximately 10^{-8} mm. Hg), ordinary air being present in the system. A blackish opaque layer resulted, the resistance of which was $30\text{ k}\Omega$.

3.1. Current-Voltage and Signal-Voltage Relations

A current-voltage curve was measured for cell A to verify that Ohm's law was obeyed. The current and voltage were measured on suitable direct reading meters. The results are plotted in Figure 1 (a), from which it is seen that Ohm's law is well obeyed for potentials of up to 70 volts across the layer, corresponding to a field of 700 v/cm. The slight deviation at higher fields is probably due to electrical heating of the layer.

With interrupted radiation falling on the layer, the relation between the a.c. signal produced and the d.c. voltage applied to the layer was determined. It was found to be linear for the same range of voltage as used in the previous experiment, as shown in Figure 1 (a).

3.2. Response Time

The response time was measured by observing the time required for the signal due to a square pulse of illumination to decay after removal of the illumination. For cell A the signal decayed to $1/e$ of its initial value in 8×10^{-4} second.

3.3. Thermo-electric Power

For both cells it was found that either junction between layer and electrode, when heated, became positive, showing the current carriers to be electrons.

3.4. Signal-Energy Relation

The relation between the signal output from the cell and the energy incident on the layer was determined using chopped radiation from a lamp of known power and temperature. The incident energy, calculated by the inverse square law, is shown plotted against the signal in Figure 1(b). For both cells the observed points lie in straight lines of slope unity, showing the signal to be proportional to the incident energy over the range of measurements, i.e. at least up to 3×10^{-3} watts.

3.5. Spectral Distribution of Sensitivity

For these measurements a lithium fluoride monochromator was used with a Nernst source, the radiation being chopped at 80 c/s. The response of the layers would be linear over the whole range of these measurements since the maximum energy used was only 10% of that used in the above signal-energy experiments. Thus the curves obtained as the ratio (cell signal)/(incident energy) are effectively constant energy spectral sensitivity curves.

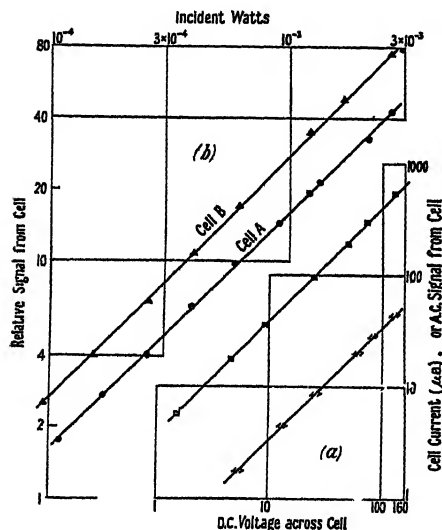


Figure 1.

(a) Current/voltage (\blacksquare) and signal/voltage (\blacktriangleright) curves. (b) Signal/energy relation.

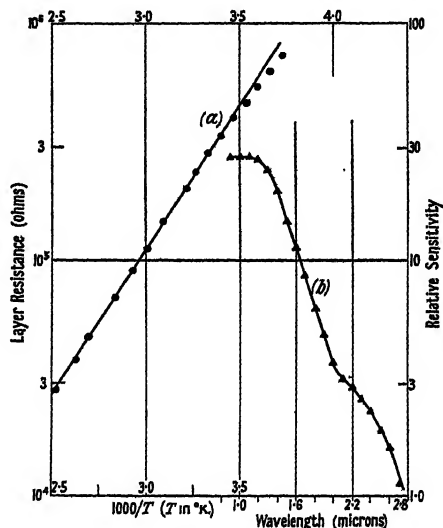


Figure 2.

(a) Resistance/temperature. (b) Spectral sensitivity.

The results obtained with cell A are shown in Figure 2(b). The curve shows a flat portion from 0.9μ to 1.2μ , then an exponential fall in sensitivity up to 2.0μ after which there is an indication of a further band of sensitivity around 2.6μ . This latter is presumably some form of impurity band, while the short wavelength band is probably due to intrinsic conductivity. For photoconductive cells there is no absolute threshold wavelength at which the sensitivity suddenly vanishes. The significant part of the spectral sensitivity curve is where the exponential fall commences. As previously (Moss 1949), the 'threshold' is taken where the sensitivity has fallen to half its value on the flat part of the curve, i.e. $\lambda_{1/2} = 1.53 \mu$.

Hall effect measurements have recently been carried out on this material by Boltaks and Zhuze (1948) from which the width of the forbidden zone is quoted as 0.6 to 0.8 eV. The figure of 0.8 eV. (which is preferred since the slope of the Hall

constant-temperature curve may be decreased by impurities) is equivalent to a wavelength of just over $1.5\ \mu$, in close agreement with the 'threshold' wavelength for the above cell.

The fact that the $2.6\ \mu$ band appears in this cell means that some 'impurities' must be present in the layer—possibly indicating that the vacuum was not sufficiently good at the time of evaporation. The cell could only be outgassed at 250°C . as higher temperatures would have involved the likelihood of evaporating the materials. However, to produce the correct rate of evaporation the tungsten heater spiral has to be run at about $1,500^\circ\text{C}$.; this would cause some evolution of gas which might well produce the observed results.

The results obtained with cell B are shown in Figure 3. The short wavelength peak corresponds to the main band of cell A, but there is now also a well defined band of sensitivity with a 'threshold' wavelength $\lambda_{1/2} = 2.6\ \mu$.

The limiting sensitivity of this cell at $1.3\ \mu$ was 2×10^{-7} watts (i.e. signal = noise in 1 c/s. bandwidth at 80 c/s.)

3.6. Resistance-Temperature Measurements

The inside of the cell was partly filled with mercury, and this could be heated electrically. Resistances were measured by battery and galvanometer, and temperatures by a thermocouple in the mercury. The results for cell A are plotted as $\log(\text{resistance})$ against reciprocal temperature in Figure 2(a). The points fall on a straight line characteristic of semiconductors, which has a slope of 0.25 ev. , indicating an activation energy of 0.5 ev. This is presumably related to the 'impurity' band of sensitivity in the $2.6\ \mu$ region. It was not possible to reach the range of intrinsic conductivity as higher temperatures caused permanent changes in the resistance of the layer.

The specific resistance at 21°C . was found to be $110\ \text{ohm cm.}$ for cell A, and $32\ \text{ohm cm.}$ for cell B, assuming a layer thickness of $1\ \mu$ in each case. The thickness of the layers was estimated from the weight of material used, the area of the layer formed, and the specific gravity, determined experimentally to be 3.4 . With the size of cell used, the layer thickness would be nearly $1\ \mu$ per milligram of Mg_3Sb_2 .

We may calculate roughly what the intrinsic conductivity would be at this temperature. Taking as the density of conducting electrons $n = 2.5 \times 10^{19} \exp(-\epsilon/2kT)$ with $\epsilon \approx 0.8\text{ ev.}$, and assuming a mobility $b = 300\text{ cm. sec}^{-1}/\text{volt cm}^{-1}$, which is the order found in recent measurements on ZnO (Hahn 1949) and silicon (Pearson and Bardeen 1949), we find $\sigma = 10^{-4}\ \Omega^{-1}\text{ cm}^{-1}$. Since the measured conductivity is about 100 times greater than this, it must be mainly due to 'impurity' levels. Hence the activation energy of 0.5 ev. found for this temperature region must be the 'impurity' activation energy, and is thus related to the $2.6\ \mu$ band of sensitivity.

For cell B the measurements were extended down to 160°K . The resulting curve of $\log(\text{resistance})$ plotted against temperature (Figure 5) is again linear over a wide temperature range, and indicates an activation energy of 0.48 ev. This is the quantum energy equivalent to a wavelength of just under $2.6\ \mu$ —which is remarkably close to the 'threshold' wavelength found from the photoconductivity curve.

It would thus appear that for this material there is little difference between the thermal and optical activation energies, although a difference would be expected in consequence of the Franck-Condon principle. From the theory put forward by

Mott and Gurney (1948), this would imply that the high and low frequency dielectric constants have approximately the same values.

The conclusions to be drawn from the measurements on these two cells are:

(a) That there is a band of intrinsic photoconductivity extending from the visible region to a 'threshold' wavelength $\lambda_{1/2} = 1.53 \mu$. The intrinsic band is not so well defined for cell B because of the large sensitivity of the 2.6μ band, but may be isolated by subtracting from the measured total the estimated form of the 2.6μ band at shorter wavelengths. Assuming that the sensitivity of this band is constant at the plateau value, or of constant quantum efficiency at the plateau value, the resultant curve in either case gives $\lambda_{1/2} \approx 1.5 \mu$. The values of the intrinsic activation energy are: optically 1.53μ and 1.5μ approximately, giving 0.81 ev., and thermally 0.8 ev. approximately, from Boltaks and Zhuze (1948).

(b) An 'impurity' band appears in both cells, but is better defined in cell B, where the photoconductivity curve gives $\lambda_{1/2} = 2.62 \mu$ (0.47 ev.). Thermal measurements on the two cells give an activation energy of 0.5 or 0.48 ev.

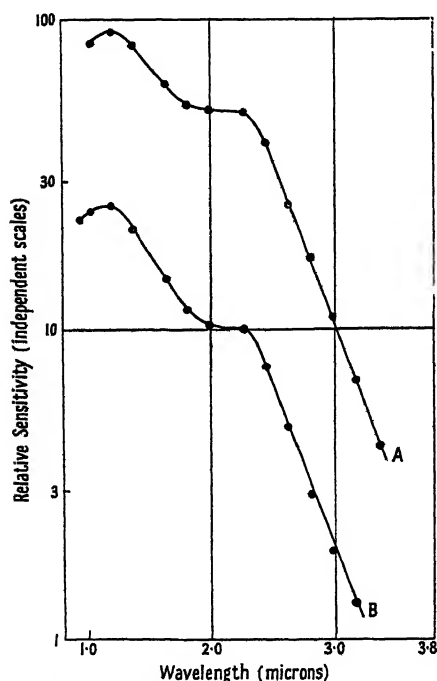


Figure 3. Spectral sensitivity of cell B. Curve A, 290°K. Curve B, 195°K.

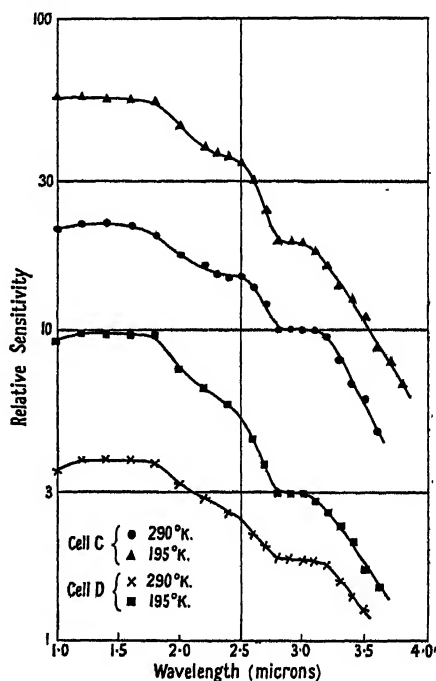


Figure 4. Spectral sensitivity curves.

It is likely that this band corresponds to the band of sensitivity found in Cd_3As_2 at 2.3μ . The wavelengths to which Zn_3As_2 and Mg_3As_2 would be sensitive would be expected to decrease progressively from 2.3μ , since the cation polarizabilities are smaller. The inference is, therefore, that antimonides are photoconductive at longer wavelengths than the corresponding arsenides.

§ 4. SPECTRAL SHIFT WITH TEMPERATURE

The two spectral sensitivity curves shown in Figure 3 were measured on cell B: at room temperature and at the temperature of solid CO_2 in acetone respectively. Both the threshold wavelengths (half the plateau value) lie at 2.62μ as read from

the curves. A repetition of the experiment gave an apparent shift with temperature of 0.02μ , which is less than the overall error of the measurements. We may say therefore that no definite shift of the threshold wavelength was observed.

From the theory put forward to explain the spectral shift in PbS, PbTe etc. (Moss 1950) it was shown that a relationship between the spectral shift, threshold energy E , and expansion coefficient β would be expected, namely,

$$\frac{dE}{dT} E^{-1/2} \beta^{-1} = \text{constant.}$$

For PbS and PbTe the constant was found to be approximately 30. Taking this value to apply to Mg_3Sb_2 , we would expect the relation $dE/dT = 30\beta E^{1/2}$.

The value of β was determined experimentally using a direct reading expansion indicator with a sample of Mg_3Sb_2 12 mm. in length. The result obtained was $\beta = 9 \times 10^{-6}$.

Hence $dE/dT = 0.018 \text{ eV}/^\circ\text{C}$. For the 95°C . change in temperature this would be equivalent to a wavelength shift of 0.09μ . Such a shift was not observed in the measurements.

Hence if the above theory of spectral shift is to apply in this case, there must be a shift of at least 0.05μ in the opposite direction, possibly of the type considered by Radkowsky (1948).

§ 5. CELLS C AND D

Two sensitive layers were prepared by evaporation in a somewhat higher pressure of air (estimated at approximately 10^{-2} mm. Hg by glow discharge). These layers had rather a black amorphous appearance.

5.1. Spectral Sensitivity Measurement

Spectral sensitivity curves were plotted for both cells at room temperature and 195°K . The curves are shown in Figure 4. All curves show the presence of an additional band of sensitivity, with a half-value threshold wavelength near to 3.5μ for cell C and 3.6μ for cell D. The other two band edges occur at somewhat longer wavelengths than in the previous cells—possibly due to a reduction in activation energies caused by higher impurity content (Heywang 1949). The conditions under which the cells were made make this seem likely.

5.2. Resistance-Temperature Measurement

Cell D was the only one made which could be cooled to liquid air temperature without causing any permanent change in the layer resistance. On cooling to 77°K . the layer resistance increased to 3×10^{10} ohms. The process was reversible and repeatable. For the electrode dimensions used and a layer 1μ thick this value corresponds to a specific resistance of 0.9×10^7 ohm cm. The specific resistances of the constituent metals at 90°K . are: Mg, 1×10^{-6} ohm cm., Sb, 10×10^{-6} ohm cm.

Hence we find that the specific resistance of the compound has increased by the remarkable factor of 10^{12} times over that of its constituents. It is to be concluded therefore that a true compound is formed, and that from the point of view of electrical properties the compound more nearly resembles an insulator than a metal.

The resistance temperature curve for cell D is shown in Figure 5. The high temperature part of the curve has an activation energy of 0.46 eV ., which is

approximately the same value as that found for cells A and B. At lower temperatures there is another linear part of the curve with an activation energy of 0.37 ev. This is not so well defined as the high temperature region, and the slope determination is not so accurate. However, a further energy level is indicated, and this would be expected to correspond to the long wavelength band of sensitivity found in these two latter cells.

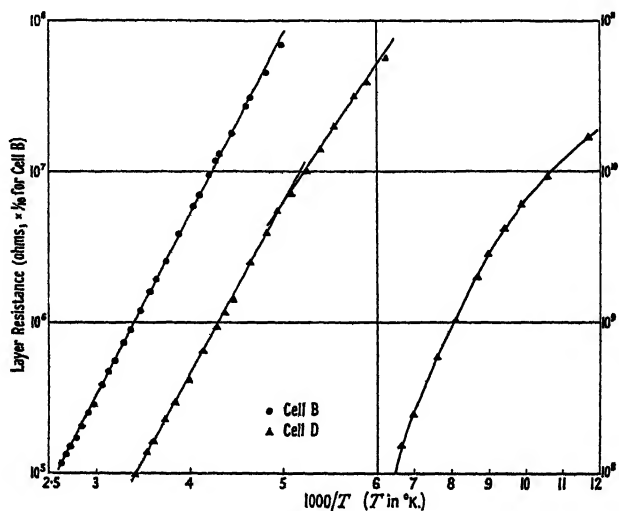


Figure 5. Resistance temperature curves.

The values of activation energy found are: optically 3.5μ and 3.6μ , giving an average 0.35 ev., and thermally 0.37 ev.. The activation energies of the three bands of sensitivity may thus be taken as 0.81 ev., 0.48 ev. and 0.35 ev.

§ 6. POSSIBLE NATURE OF THE IMPURITIES

The term 'impurity' is used in a very wide sense in this discussion, and does not necessarily refer to foreign atoms. Any lattice imperfections might produce energy levels in the forbidden zone and lead to the results observed. The materials used to make the layers were of high purity, and care was taken to prevent contamination during preparation. Hence the only foreign atoms likely to occur in the layers are those from the gas present during the evaporation, i.e. mainly oxygen and nitrogen. However, the gas may not necessarily combine with the layer, but may merely cause the layer to deposit in a less regular form.

The impurity levels may well be due to a stoichiometric excess of one of the constituents, the excess appearing in the lattice interstitially or as a consequence of vacant lattice sites of the other constituent.

§ 7. HARDNESS

It is interesting to note that the compound Mg_3Sb_2 is much harder than its constituents. The compound would just scratch natural quartz, indicating a hardness figure close to 7, compared with 2 for Mg and 3 for Sb (*Handbook of Chemistry and Physics*).

§ 8. CONCLUSIONS

Photosensitive films of Mg_3Sb_2 have been prepared by vacuum evaporation. At low light levels the photo-response is proportional to the incident light intensity.

Layers made in high vacuum show one main band of photoconductivity, with a long wavelength 'threshold' at 1.5μ or 0.8 ev. This corresponds closely to the value of intrinsic activation energy found from Hall effect measurements by other workers.

When the layers are evaporated in a low pressure of air, two other bands of sensitivity appear. These have threshold wavelengths at 2.6μ (0.47 ev.) and 3.5μ (0.35 ev.) respectively. Resistance temperature measurements give activation energies of 0.48 ev. and 0.37 ev. The thermal and optical activation energies are thus nearly equal for all the bands.

No appreciable shift of the threshold wavelength was observed on cooling to 195°K .

Since the work was concluded, a brief abstract has appeared stating that the Russian workers Zhuze *et al.* (1948) have observed photoconductivity in Mg_3Sb_2 .

ACKNOWLEDGMENT

Acknowledgment is made to the Chief Scientist, Ministry of Supply, for permission to publish this paper.

REFERENCES

- BOLTAKS, B. I., and ZHUZE, V. P., 1948, *J. Tech. Phys. U.S.S.R.*, **18**, 1459 (abstract in *Science Abstracts A*, 1949, **52**, 1904).
HAHN, E. E., 1949, *Phys. Rev.*, **75**, 1631.
Handbook of Chemistry and Physics, 1948 (Ohio: Chemical Rubber Publishing Co.).
HANSEN, M., 1936, *Aufbau der Zweistofflegierungen* (Berlin: Springer).
HEYWANG, W., 1949, *Z. Naturforschung*, **4a**, 654.
Landolt-Bornstein Physical Chemical Tables, 1936 (Berlin: Springer).
MOSS, T. S., 1949, *Proc. Phys. Soc. B*, **62**, 741; 1950, *Ibid.*, **63**, 167.
MOTT, N. F., and GURNEY, R. W., 1948, *Electronic Processes in Ionic Crystals* (Oxford: Clarendon Press).
PAULING, L., 1927 a, *Proc. Roy. Soc. A*, **114**, 181; 1927 b, *J. Amer. Chem. Soc.*, **49**, 765.
PEARSON, G. L., and BARDEEN, J., 1949, *Phys. Rev.*, **75**, 865.
RADKOWSKY, A., 1948, *Phys. Rev.*, **73**, 749.
ZHUZE, V. P., MOCHAN, I. V., and RUIRKINS, S. M., 1948, *J. Tech. Phys. U.S.S.R.*, **18**, 1494 (abstract in *Science Abstracts A*, 1949, **52**, 7013).

The Rehbinder Effect

By E. N. DA C. ANDRADE,* R. F. Y. RANDALL† AND

M. J. MAKIN‡

University College, London

MS. received 2nd August 1950

ABSTRACT. According to Rehbinder the immersion of wires of certain metals in non-polar paraffin containing a little oleic acid increases the rate of flow of the metal under a given stress and increases the electrical resistivity. The experiments described show that the mechanical effect can be obtained with single crystals of cadmium if the surface is contaminated by a thin oxide layer, which is known to increase the critical shear stress, but not if the surface is clean. The effect is accordingly attributed to the disruption of the hardening surface layer by the active agent and not to penetration into the metal, as assumed by Rehbinder. It has not been found possible to detect any electrical effects with cadmium or lead crystal wires having either clean or contaminated surfaces.

REHBINDER and his collaborators have described in a series of publications (1931, 1937, 1941) striking effects of a surface active liquid, present in small quantities in a non-polar hydrocarbon liquid, on the mechanical and electrical properties of metals and other substances. The contact of such liquids with polycrystalline specimens of lead, tin and copper is said to accelerate plastic flow under a given stress, and to produce an increase in the electrical resistivity of the metal. Particularly striking results are recorded with single crystals of tin and zinc, of diameter 0.2 to 2 mm., extended in a solution of 0.2 % oleic acid in a non-polar hydrocarbon oil, 0.2 % having been found to be an optimum concentration. This liquid will be called 'Rehbinder's reagent' in this paper. The yield stress in a particular case is recorded as reduced from 265 to 130 gm.wt/mm², and the stress needed to maintain a constant rate of extension as reduced by the liquid to the extent of some 50 %, over a range of elongations. Marked changes in electrical resistance were also recorded. In one case the resistance of a crystal extended by 180 % in the liquid was four times that of a similar crystal extended in air, the resistance being measured while the stress necessary to produce the extension was still applied. Rehbinder further found that on removal of the stress the resistance of the crystal slowly diminished, tending towards its normal value. He attributed all these effects to the penetration of thin films of the active medium into minute cracks developed during the deformation. The increase of electrical resistance, for instance, was attributed to the insulating action of the films of active material, the recovery on relaxation of load being due to the gradual healing of the cracks, with expulsion of the liquid films.

Recently D. S. Kemsley (1949) published a brief account of experiments on single crystals of tin, which showed no effect of Rehbinder's reagent on the critical shear stress or on the electrical resistance.

Work in this laboratory has shown the great effect which oxide, or similar, films on the surface of a metal single crystal have on its mechanical properties. Roscoe (1936) first drew attention to the marked effect of oxide films in raising the critical shear stress. Recent experiments on which a preliminary note has

* Now Director, The Royal Institution, London.

† Now at the University of Sheffield.

‡ Now at the Davy Faraday Research Laboratory, The Royal Institution, London

appeared (Andrade and Randall 1948) have shown that the contact of certain electrolytes with the surface of cadmium single crystals produces effects which have been traced to surface layers. Crystals loaded to flow at a slow rate have shown an immediate increase in rate of flow of roughly three times when immersed in a solution containing cadmium ions, whereas crystals of which the surface has been thoroughly cleaned by thermal evaporation *in vacuo* show no increase in rate on immersion. These experiments made it seem possible that the mechanical effects observed by Reh binder might be due to a disintegrating action of the surface agent on such films rather than to a deep penetration of the agent into minute cracks (Andrade and Randall 1949). The following experiments were carried out to investigate this point.

The crystals used were made from clean spectroscopically pure cadmium wire, 1 mm. diameter, by the Andrade and Roscoe method (1937). The disposition of the crystal axes was found by standard x-ray back-reflection methods: in the diagrams ψ indicates the angle between the glide plane and the wire axis and λ the angle between the wire axis and glide direction.

The measurements were made with apparatus already described which permitted the wire, slowly extending under a fixed load, to be surrounded by a chosen liquid at any desired moment. Any weakening effect due to the slow rupture or disintegration of the oxide film would be expected to show as an increase in the rate of flow.

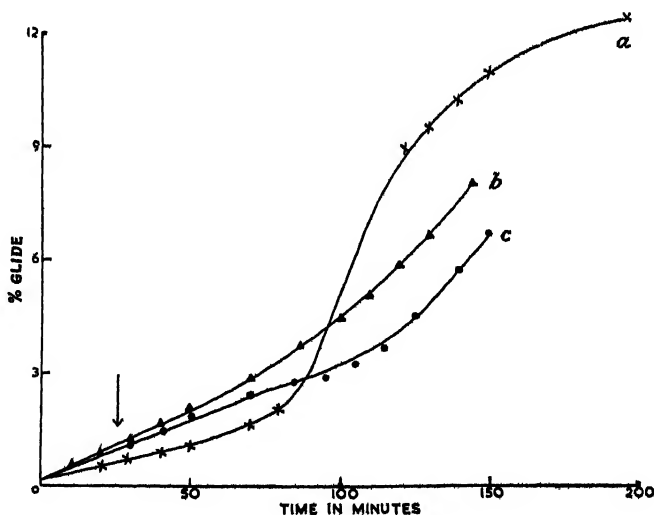


Figure 1. Single crystals of cadmium.

0.2% Oleic acid in paraffin applied at time indicated by arrow: (a) oxidized by long exposure to the atmosphere; (b) oxidized by heating in oxygen and cleaned; (c) oxidized by heating in oxygen and not cleaned.

Resolved shear stress: (a) 27.4 gm.wt/mm², (b) 19.1 gm.wt/mm², (c) 24.9 gm.wt/mm²; $\psi=39^\circ$; $\lambda=43^\circ$.

Figure 1 shows the results of the oleic acid mixture on three lengths of crystal, all from the same specimen: (a) is a crystal which had been left lying about for some time in air. The surface looked dirty, and no doubt was covered with an oxide film, but this was not visible as a coloration; (b) is a crystal which had been covered with a yellow film of oxide by heating in air at about 250° C. for four hours. It was subsequently cleaned by washing with N/10 HCl, water, N/10 NaOH and

then water, the method adopted by Rehbinder; (c) is a similar crystal which had not been cleaned. The resolved shear stresses on the glide plane and in the glide direction were 25, 18 and 23 gm.wt/mm² respectively.

After about half an hour's immersion, the rate of flow of (a) started to increase and eventually reached a value some twelve times the initial rate. About 100 minutes after the introduction of Rehbinder's reagent, hardening began, as normally occurs from the start with a wire of this inclination.

Wire (b), rapidly oxidized and cleaned by Rehbinder's method, was the softest. It showed an increase of rate after half an hour, increasing to about 2.5 times the initial rate. Wire (c), not cleaned, was a little harder and showed a slightly more marked acceleration effect, the rate increasing to about 3.2 times. It thus appears that the light, but very slowly formed, oxide layer which is built up by exposure to the normal atmosphere produces the greatest mechanical effects.

Figure 2 refers to a crystal whose surface has been cleaned by thermal evaporation, the wire having been heated *in vacuo* by the passage of an electric current until marked deposition of vaporized cadmium was visible on the walls of the containing glass tube (Andrade and Randall 1950). As can be seen, there is no change of rate after the introduction of the liquid except the normal slight diminution—a result in accordance with the supposition that the Rehbinder effect is due to a surface layer of oxide.

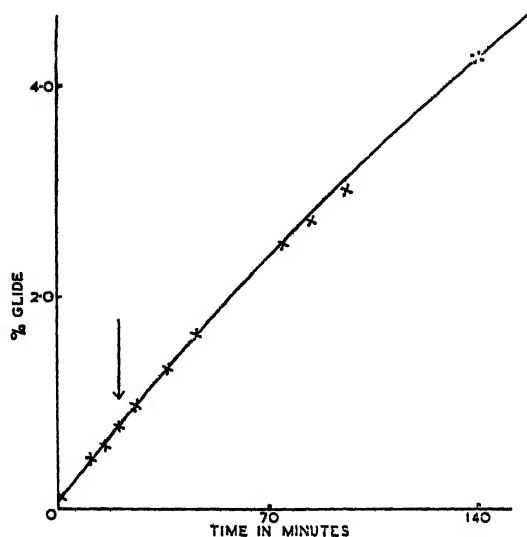


Figure 2. Single crystal of cadmium.

0.2% Oleic acid in paraffin applied at time indicated by arrow. Specimen with surface evaporated *in vacuo*. Resolved shear stress = 30 gm.wt/mm²; $\psi = 21^\circ$; $\lambda = 30^\circ$.

Attention was then directed to the electrical effect. In the absence of precise knowledge as to how Rehbinder measured the electrical resistance it seemed possible that the large increase recorded by him might have been caused by the penetration of a film of the active agent between the grips holding the ends of the specimen and the wire itself.

Preliminary experiments were tried with simple apparatus in which the fully extended (100–200 %) crystal was held under tension by springs. The resistance

before and after immersion in the Rehbinder's reagent was determined by measuring with a potentiometer the difference of potential between the grips when a steady current was passing through the specimen. No significant increase in resistance could be detected in any experiment, although tests were made on normally prepared crystals of cadmium, oxide-coated crystals of cadmium, and polycrystalline cadmium. Tests were also carried out on oxide-coated lead crystals.

Typical sets of results are given below for oxide-coated cadmium and lead crystals, the times being measured from the introduction of the reagent, so that negative times are before the introduction.

Oxide-coated Cadmium Crystal									
Time (min.)	...	-40	-20	20	40	70	160	260	2½ days
Resistance*	...	2.33	2.33	2.33	2.34	2.35	2.34	2.34	2.44

Oxide-coated Lead Crystal									
Time (min.)	...	-22	-2	18	38	58	78	93	1 day
Resistance*	...	0.829	0.819	0.819	0.817	0.817	0.816	0.815	0.804

* in arbitrary units.

In view of the complete failure to reproduce Rehbinder's results it was decided to perform mechanical and electrical measurements simultaneously and the apparatus previously used was modified for this purpose.

The disposition was as shown in Figure 3. The only parts not of metal are the two shaded columns and insulating plug, which were of ebonite. W is the cadmium

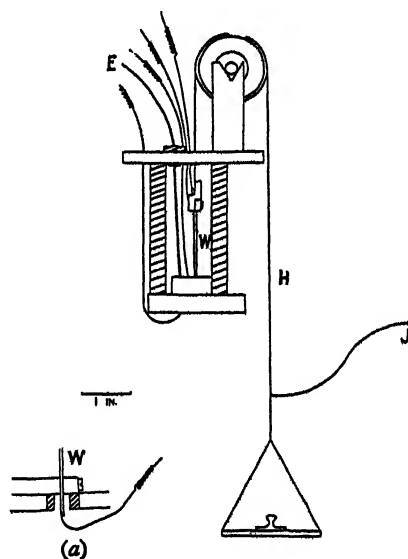


Figure 3.

wire, the load being applied by means of the flexible metal tape H passing over the flat pulley. The connections for the current were the wire E and the flexible tape leading from H to J. Potential leads were soldered to the upper and lower chucks

and also to the ends of the specimens, which passed right through the chucks as shown in inset Figure 3(a). Using these leads, changes in potential difference between the ends of the specimen and between the chucks could be measured. The results were expressed as the ratio of these potential differences to the potential difference across a short length of unstrained polycrystalline cadmium wire in series with the specimen: this obviates the necessity for keeping the current constant over long periods of time. In this way changes in resistance of the specimen could be followed and any spurious resistance at the chuck detected. A steady direct current of one ampere was passed through the specimen and the potentials of the various leads measured in rapid succession using a commercial potentiometer reading to 1 microvolt. Successive results were found to differ by less than 5%. No special attempt was made to control the temperature of the apparatus as the magnitude of the effect found by Rehbinder was too large to make this necessary.

While the electrical measurements were in progress the increase of length of the specimen was measured with a travelling microscope.

The results are shown in Figure 4. No Rehbinder effect was observed with a cadmium single crystal prepared *in vacuo* and tested immediately on exposure to the air (curve (a)). After oxidation by heating in oxygen for four hours at 200° c. a length of the same crystal gave the curve (b) which shows the appreciable hardening due to the oxide and also an increase in the rate of flow rising to 3.3 times the normal rate and commencing some 100 minutes after immersion in the liquid. The slow rise in the electrical resistance of the specimen is entirely accounted for by the change in shape as the extension proceeds. No effect due to the possible penetration of the active agent into the chucks could be detected. Other cases were obtained in which crystals showed a like change in rate of extension without change in resistivity.

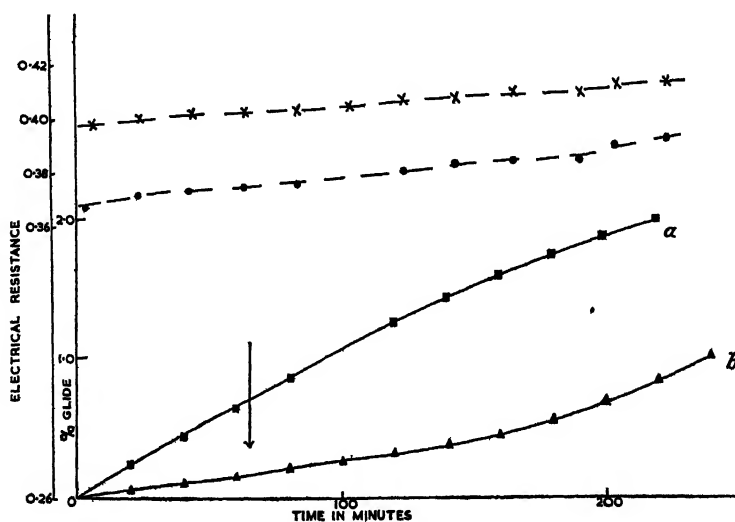


Figure 4. Single crystals of cadmium.

0.2% oleic acid in paraffin applied at time indicated by arrow; (a) glide for specimen tested immediately on exposure to the air; (b) glide for specimen tested after oxidation.
 • Electrical resistance for specimen giving glide curve (a); x electrical resistance for specimen giving glide curve (b). Resolved shear stress = 33.2 gm.wt/mm²; $\psi = 25^\circ$; $\lambda = 31^\circ$.

It is clear therefore, that the effects of surface active agents on the mechanical behaviour can occur without any change in resistivity. The fact that we have failed to obtain any marked changes in resistivity suggests that the changes recorded by Rehbinder were associated with some peculiar experimental conditions not essential to the mechanical effect. We conclude that the behaviour of metals in the Rehbinder reagent is due to a disintegration of a surface film and not to deep penetration of the surface active component into minute cracks.

After the work here described was practically completed a paper by S. Harper and A. H. Cottrell appeared (1950) in which experiments were described indicating that the Rehbinder effect is attributable to the action of the reagent on an oxide film. These authors bring strong support to this view by showing a direct proportionality between the viscosity of the reagent and the time taken to produce the effect. They used zinc crystals, with which the effect took place much more rapidly than with our cadmium crystals, in seconds where we found minutes. In view of these completely independent results with two different metals there seems little doubt that the Rehbinder effect as far as mechanical properties of metals is concerned is due to a disintegrating effect on oxide films. Harper and Cottrell made no attempt to detect an electrical effect.

ACKNOWLEDGMENTS

Our best thanks are due to Mr. R. King for helpful discussions during the experiments. Dr. Randall was in receipt of a grant from the Ministry of Supply at the time this work was carried out and Mr. Makin has been helped by a grant from the Department of Scientific and Industrial Research, both of which are gratefully acknowledged.

REFERENCES

- ANDRADE, E. N. DA C., and RANDALL, R. F. Y., 1948, *Nature, Lond.*, **162**, 890; 1949, *Ibid.*, **164**, 1127; 1950, *Proc. Phys. Soc. B*, **63**, 198.
ANDRADE, E. N. DA C., and ROSCOE, R., 1937, *Proc. Phys. Soc.*, **49**, 152.
HARPER, S., and COTTRELL, A. H., 1950, *Proc. Phys. Soc. B*, **63**, 331.
KEMSLEY, D. S., 1949, *Nature, Lond.*, **163**, 404.
REHBINDER, P., 1931, *Z. Phys.*, **72**, 191; 1947, *Nature, Lond.*, **159**, 866.
REHBINDER, P., LICHTMAN, V. I., and MASLENNIKOV, V. M., 1941, *C.R. Acad. Sci., U.R.S.S.*, **32**, 125.
REHBINDER, P., and LOGGINOV, C. I., 1941, *C.R. Acad. Sci., U.R.S.S.*, **30**, 491.
REHBINDER, P., and WEINSTRÖM, E. K., 1937, *Bull. Acad. Sci., U.R.S.S.*, No. 4/5, 531.
ROSCOE, R., 1936, *Phil. Mag.*, **21**, 339.

A Study of the Magneto-Resistance of Silicon-Iron

By R. PARKER

University of Nottingham

Communicated by L. F. Bates; MS. received 18th May 1950

ABSTRACT. Iron containing a small percentage of silicon in solid solution has an unusually small magneto-resistance. Longitudinal and transverse curves of $\Delta\rho/\rho$ against field strength were obtained for a number of single and polycrystalline specimens. In large fields the curves show a linear decrease, which is less than 0.02 times that of nickel and cobalt in the corresponding field range. In small fields various unusual features are found. It is concluded that the small magneto-resistance of polycrystalline silicon-iron is due to the small contribution of individual crystallites and not to an averaging-out process.

§ 1. INTRODUCTION

THE magneto-resistance, i.e. the change in electrical resistance on the application of a magnetic field, of the ferromagnetic metals and their alloys, can be divided into two distinct parts: a large decrease in resistivity in the neighbourhood of the Curie point owing to an increase in spontaneous magnetization, and a much smaller, more complicated, change associated with technical magnetization curves. The first effect has been studied in detail only in nickel by Potter (1931) and Gerlach (1932) and was interpreted theoretically by Mott (1936). The second effect has been investigated by well over a hundred workers, though single crystal investigations have been reported only by Kaya (1927) and Döring (1938) in nickel, and by Webster (1926, 1927) and Shirakawa (1940) in iron.

Their main results can be described in the following way: the ratio $(\Delta\rho/\rho)_{||}$, i.e. the relative change in electrical resistivity on magnetizing a specimen parallel to the direction of current flow, is increased steadily from zero to a constant final value; on transverse magnetization (perpendicular to the current direction), a steady decrease usually takes place. Both values become practically constant when technical saturation is achieved, and further changes take place only when the applied field increases the spontaneous magnetization appreciably. The technical saturation value of $(\Delta\rho/\rho)_{||}$ for iron is approximately 3×10^{-3} but depends to a considerable extent on temperature, purity and previous heat treatment of the specimen. The $(\Delta\rho/\rho)_{\perp}$ curves for nickel and iron are precisely those one would expect from a large aggregate of randomly orientated single crystals. From the simplicity of the magneto-resistance curves and considerations of cubic symmetry, Döring derived a general relation for $\Delta\rho/\rho$ in terms of five constants $K_1 \dots K_5$, and even powers of α and β , which represent the direction cosines of the magnetization vectors and the current direction respectively, referred to the main crystal axes. It will be shown later that this analysis is not applicable to silicon-iron alloys of certain compositions.

Shirakawa (1936-39) made a study of the longitudinal magneto-resistance of a large number of ferromagnetic alloys as a function of temperature, among them silicon-iron alloys of certain percentages. From his results it is concluded that the saturation value of $(\Delta\rho/\rho)_{||}$ at room temperature changes from positive to

negative as the silicon content exceeds 1.7%. Further, it appears that this value is much smaller in the region of 1 to 3% silicon than that of any other ferromagnetic material hitherto measured. This has been confirmed by Drozhima and Shur (1947). Before any explanation of this phenomenon can be offered, it is necessary to investigate whether the small coefficient is due to a genuine reduction in the contribution of individual crystallites. This was one reason for undertaking the present investigation. Owing to the smallness of $\Delta\rho/\rho$ a very delicate measuring technique is required to record accurately changes in resistance of 10^{-7} ohm in the specimen under test.

§ 2. EXPERIMENTAL PROCEDURE

The method of measurement employed was governed by the facts that single crystal specimens available were usually in strip or rod form, 2–4 cm. in length, that their magneto-resistance coefficient is smaller by a factor of 10 than that of most other ferromagnetic metals, and that the cross-sectional area of rods could not be reduced below 2 mm² without introducing appreciable strains on mounting, as well as increasing the magnetic field due to the specimen current to an undesirable value. The actual resistance of specimens was therefore sometimes less than 10^{-2} ohm, so that a simple potentiometer or Kelvin double bridge circuit was out of the question. The principle of the method here used has already been described by Bates (1946) in an investigation on high coercivity alloys. It has been found suitable, with only minor modifications, to achieve the extra sensitivity required and is therefore only described very briefly here.

Roughly the principle is as follows: a constant current of approximately 1.018 amp. is sent through the specimen, across which a potential difference of the order of 10^{-2} volt is established. The potential difference is balanced, via a subsidiary circuit, against that produced by an accumulator across a suitable potentiometer. A sensitive galvanometer then measures the out-of-balance current when the specimen is placed in a magnetic field. The instrument used for most of the results described was a Tinsley galvanometer, which gave a deflection of nearly 2 mm. at 1.85 metres scale distance, due to a change of 10^{-8} volt in the circuit, when used on the most sensitive range. It will be seen, therefore, that magneto-resistance coefficients of 10^{-6} are detectable.

The specimen holders used for rods were similar to those described by Bates. For strips a special design was used to give minimum transverse demagnetizing factors and to avoid putting strains on to the specimens when the potential leads were screwed on. It was decided not to solder the latter to avoid possible chemical complications. The circuit was readily calibrated to the desired accuracy.

All silicon-iron specimens were well annealed at 800–850° C. *in vacuo* just before use. It was ascertained for each specimen separately that the process of mounting it did not affect the results significantly owing to unavoidable strains. The magnetic field was supplied by an electromagnet and was measured with a magnetic potentiometer in the longitudinal, and by a search coil in the transverse experiments. Both instruments were previously calibrated in a standard solenoid. The potentiometer measures directly the field acting inside the specimen. In the case of the transverse measurements, a rough correction for the demagnetizing factor could be made if desired, although the uncorrected, applied field is always plotted here. Before each reading was taken, the specimen was carefully demagnetized.

A number of experimental difficulties had to be overcome. To avoid thermoelectric disturbances, the number of junctions of unlike metals was reduced to a minimum and these were well lagged. The specimen had to be shielded from draughts owing to the large temperature coefficient of resistivity of iron. A simple calculation shows that a change of temperature of 0.02°C . would cause a $\Delta\rho/\rho$ of the same magnitude as the magneto-resistance measured in silicon-iron. This was not serious, as a constant temperature had to be maintained only while a reading was being taken, and the time, therefore, could be reduced to approximately 15 seconds when the stray induction pick-up in the galvanometer circuit was counteracted by a compensating coil; but it limited the maximum field that could be applied to specimens to about 10,000 oersteds. Larger magnetizing currents produced objectionable heating effects even in the short time specified. The most serious difficulty was the avoidance of leaks from the magnet (200-volt) supply to the measuring circuit. The resistance between the two had to be maintained at 10^{12} ohms and sensitive experiments could not be carried out successfully on days of high humidity.

Before making experiments with the above specimens the sensitivity of the method was tested by preparing a specimen of copper, with a calculated resistance of 0.003 ohm. The change of resistance expected was of the order of 5×10^{-8} ohm on applying a field of 10,000 oersteds. This was readily detected above the unavoidable deflections due to leaks, and the results were consistent with those quoted for this field in the Chemical Rubber Company's *Handbook of Physics and Chemistry*.

§ 3. RESULTS AND DISCUSSION

It was verified that, for a polycrystalline and for two single crystal specimens $\Delta\rho/\rho$ did not change (within the limits of experimental error) upon halving or reversing the current through them. Both results are consequences of Ohm's law and have been confirmed for other metals by Webster (1926) and Stierstadt (1930). It was also found that the remanent magneto-resistance was very small, and less than 5% of the saturation change in single crystal specimens. No detectable difference could be observed in the subsequent magneto-resistance, when a sample was brought to the remanent state by a few large or many small current steps. This was put down to the great magnetic softness of silicon-iron when well annealed. The statement was not found to hold for unannealed nickel strips.

Although a smooth ($\Delta\rho/\rho$, H) curve could always be obtained as H was increased, small spontaneous variations of $\pm 10\%$ were observed in low fields, and of about half this amount in the saturation value, with single crystals. Variations in the case of polycrystals were considerably smaller; they occurred usually, but not entirely, after resoldering or setting the specimen, and could not be improved by successive annealing. They are undoubtedly due to changes in internal strain, which contribute an energy term which easily swamps the small magneto-crystalline energy of silicon-iron. In view of the considerable variations reported in measurements on 'identical' single crystals, it can be concluded that the present measurements show smaller variations than would be expected, if the experiments were repeated with different specimens cut from the same material in the stated directions. The difference between $(\Delta\rho/\rho)_{\parallel}$ and $(\Delta\rho/\rho)_{\perp}$ at saturation is, of course, not affected. A full discussion of these features is given by McKeehan

(1930) and Becker and Döring (1939). It should be noted that the shape of the curves is always the same.

Polycrystalline Specimen (1.8% Si). Polycrystalline specimens were cut from a sheet of rolled metal which was supplied as 'genuinely polycrystalline'. In view of the possibility of a preferred crystal orientation, two specimens cut perpendicular to each other were examined and it was found that the differences were insignificant. The shapes of the curves were almost exactly the same when the demagnetizing factor was allowed for, a typical curve being recorded in Figure 1. The following

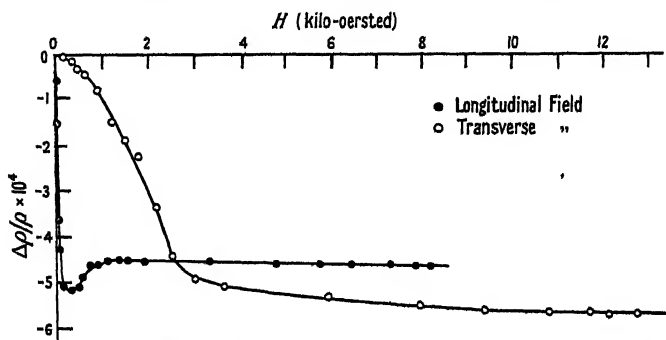


Figure 1. The magneto-resistance of 1.8% silicon-iron, polycrystalline.

features of interest may be noted. There is a small but definite linear decrease in ρ in the region of increase in spontaneous magnetization. It is about 50 times smaller than that for the corresponding field range in nickel (Englert 1932) or cobalt (Bates 1946). It is very likely that this decrease would also occur in pure iron, but no data of sufficient accuracy are available.

The saturation value of $(\Delta\rho/\rho)_{||}$ fits well into the scheme arising from Shirakawa's results for 1.66% and 5.18% silicon-iron at room temperature. The next point of interest is the small but marked minimum in the $(\Delta\rho/\rho)_{||}$ curve in the region of 200–400 oersteds, which could be detected in both specimens measured. No other material well below its Curie point shows such complexity. Its significance will be referred to later on. Another unusual feature is the small value of 'B' (using Döring's notation), which denotes $(\Delta\rho/\rho)_{||} - (\Delta\rho/\rho)_{\perp}$ at saturation. The only other comparable value is that reported for alnico by Bates (1946). Finally, the relation between longitudinal magnetostriction $\Delta l/l$ and magneto-resistance coefficients is of interest. It was thought for some time that the two phenomena were connected by a simple relation. It has, however, been shown that in a polycrystal a simple relation is demanded by the fact that both phenomena are simple functions of the magnetization vectors with respect to the crystal axes (Becker and Döring 1939). That this relation is of no more fundamental nature is demonstrated by comparing Figure 1 with the magnetostriction curves for the same specimen, Figure 2. The actual value of the magnetostriction coefficient is uncertain to within 30% owing to the unfavourably shaped specimens available, but the shape of the curve, and its variation with tension, is exactly reproducible. The addition of 1.8% silicon has not appreciably changed the magnetostriction properties from those of iron. The magneto-resistance curve, however, is different in shape and magnitude. It is to be expected, therefore, that on plotting $(\Delta\rho/\rho)_{||}$ against $\Delta l/l$ Figure 3, a curve quite different from that for pure iron (Kornetzki 1943), is obtained. It will be noted, however, that in

the region of the magnetization curve where rotations from the easy directions do not occur, the two phenomena are proportional to each other, as required from symmetry considerations (Becker and Döring 1939).

Single Crystal Strip cut along a [110] Direction, in a (001) Plane (2.8% Si). Actually, the strip plane of this specimen was inclined at approximately 4° to the (001) plane. Results on this crystal are of great importance, for we believe that

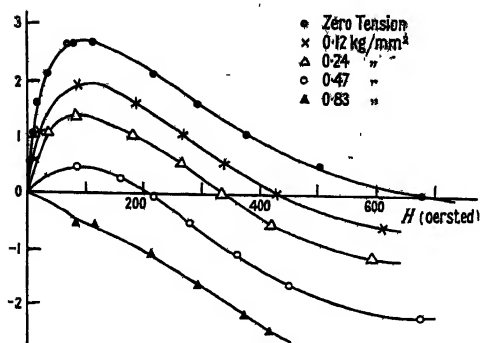


Figure 2. Longitudinal magnetostriction curves of 1.8% silicon-iron, polycrystalline.

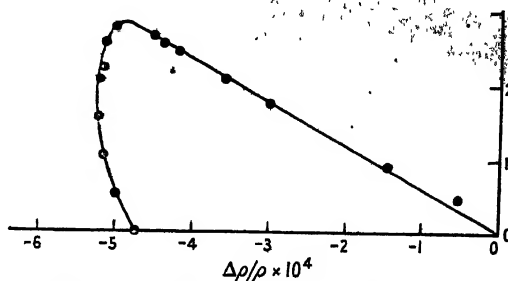


Figure 3. Relation between magneto-resistance and magnetostriction in 1.8% silicon-iron.

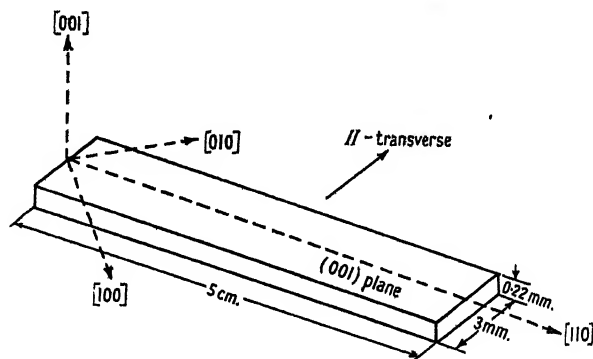


Figure 4. Diagram of single crystal strip of 2.8% silicon-iron.

we now have a reliable picture of the domain processes taking place on magnetization, based on Néel's treatment of the problem. This particular specimen was studied extensively by means of powder photographs by Bates and Neale (1949, 1950) and found to conform to Néel's theoretical predictions. A diagram of the specimen and the magneto-resistance curves obtained are given in Figures 4, 5 and 5 (a) respectively.

Apart from the small but definite difference in slope between the longitudinal and transverse curves in very large fields, which is also apparent in Kaya's results on nickel single crystals, the most interesting feature is the complicated behaviour in longitudinal fields below saturation. The initial minimum in the curve is certainly not due to experimental error. It was obtained after repeated annealing, resoldering etc. The possibility of faulty alignment with the field was ruled out by varying the specimen axis by 3° with respect to the applied field, when no appreciable change in the curve resulted. Even if the specimen were turned through a larger angle, the contribution of the transverse effect would here be small, owing to the large demagnetizing factor. It was therefore concluded that

the complex curve represents the genuine magneto-resistance contribution. The actual variations in the minimum value were found in a region where considerable variations in the Bitter figure line spacing were observed and must be put down to unavoidable small strains. No curve for other single crystals has shown such complexity, and it is even more surprising when we examine the simplicity of the magnetization process. Apart from 180-degree boundary displacements in the region $H < 20$ oersteds, only domain rotations from the [010] and [100] directions into the [110] direction take place; from considerations of crystal symmetry both types of rotation contribute equally to the magneto-resistance. (The [001] direction is ruled out by shape anisotropy.) The change

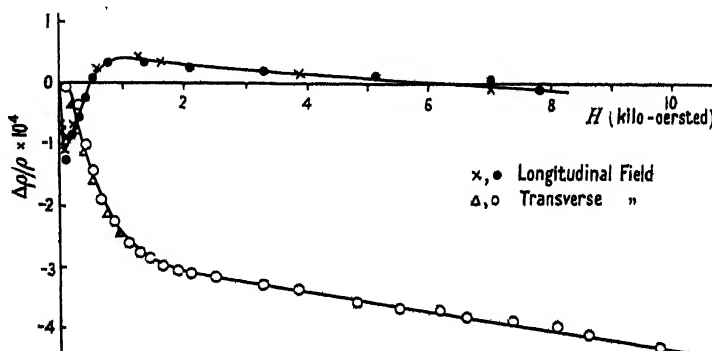


Figure 5. Magneto-resistance curves for single crystal strip of Figure 4, 2 sets of independent measurements.

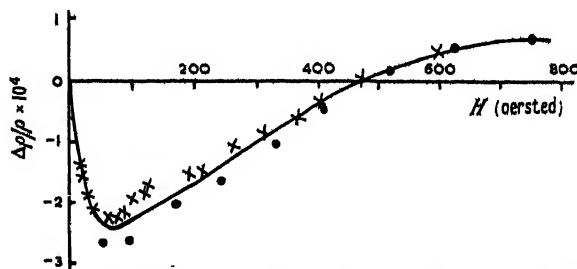


Figure 5a. Part of the longitudinal curve of Figure 5, enlarged.

in resistivity is too complex to be described by Döring's approximation $\Delta\rho = f(\alpha^2, \alpha^4)$, so that for silicon-iron his whole analysis must break down. This is quite apart from the fact that in his equation the constant $\frac{1}{2}$, introduced to allow for the random distribution in the demagnetized state, must be replaced by a function of α and β , to allow for the shape anisotropy.

Two conceivable explanations of this curve are found untenable on close examination. On Néel's theory, certain closure domains on the side of the strip are magnetized parallel to the [001] direction. If changes of the magnetization vectors from [001] to [110] and from [100] to [110] are associated with a decrease and increase in resistivity, respectively, the two opposing effects might produce a minimum. Both from experimental and theoretical considerations we conclude that the volume of such closure domains does not exceed 0.1% of the total volume of the specimen. To make comparable contributions, the magneto-resistance associated with such domains must be more than 500 times

that of the other magnetization processes. From Figure 5 it is readily calculated that this would imply a contribution to $(\Delta\rho/\rho)_{||}$ of 0.02, which is several times greater than that for pure iron.

A more reasonable explanation would seem to connect the minimum of the curve with the number of Bloch walls traversed by the current. Bitter figure experiments showed that with the specimen used here, the line spacing at 80 oersteds is approximately 100μ . Though a line pattern cannot be obtained in the demagnetized state, the number of lines is very much smaller for a very weak field, and the spacing is certainly over 250μ . This is in reasonable agreement with similar patterns obtained by Williams, Bozorth and Shockley (1949). Since each line indicates the presence of two Bloch walls when the strip surface is not a perfect crystal plane, it is concluded that the number of walls increases from 80 to 200 per cm. in the region of decreasing $\Delta\rho/\rho$. As the effective specimen length was nearly 4 cm., a total of some 480 walls appear when a field of 80 oersteds is applied. The resistance of the specimen between the potential leads was found to be 0.025 ohm by an independent experiment. The appearance of 480 walls would therefore cause a decrease in resistance of $0.025 \times (\Delta\rho/\rho)_{80 \text{ oe}} = 6.25 \times 10^{-6}$ ohm, and the decrease associated with 1 wall $\approx 1.3 \times 10^{-8}$ ohm.

Now, is it very difficult to see why this should be the case. A 90-degree Bloch wall must be looked upon as a region of greater spin randomness than a corresponding region inside a domain and might be expected to contribute an

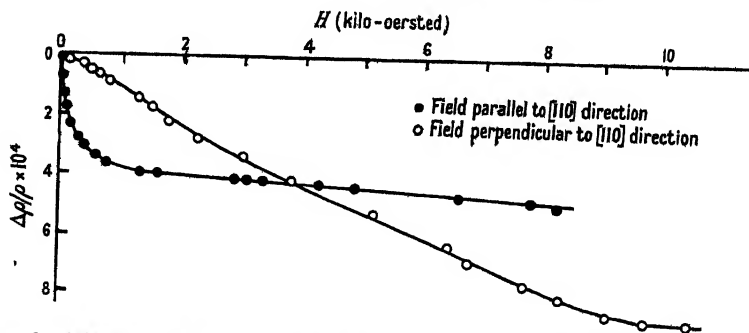


Figure 6. Magneto-resistance of a single crystal strip, specimen surface in (110) plane.

increase rather than a decrease in resistance (Mott 1936). Moreover, if we take the length of a unit cell along a $[110]$ direction as 4 Å. and the effective thickness of the domain wall as extending over 60 unit cells, we find that the resistivity could be reduced to zero, if a state in the metal could be created so that it consisted entirely of walls.

We are led to the conclusion, therefore, that the minimum is due to a rotation of the spontaneous magnetization vectors from a position of 45° to the field into one making a smaller angle. It is reasonable to believe that the slight minimum in the polycrystalline curve is due to the contributions from crystallites of this type. It is hoped, shortly, to discuss elsewhere the reason for this unusually complex behaviour from the point of view of the theory of electrical conductivity in metals.

Specimen Axis $[110]$, Plane (110), (2.6% Si). A specimen was cut along a $[110]$ direction with a rectangular cross section, the longer side of which was contained in a (110) plane. The specimen had an unavoidably large transverse demagnetizing factor, which is demonstrated by the fact that $(\Delta\rho/\rho)_\perp$ required an applied field of 10,000 oersteds to reach approximate saturation (Figure 6). The

general shape of the curves is not unusual; the actual coefficients are much smaller than for pure iron.

Specimen Axis [100], *Plane* (110). (2.6% Si). A diagram of the specimen and the curves obtained with it are recorded in Figures 7 and 8 respectively. This specimen also had a large transverse demagnetizing factor, and it is possible that $(\Delta\rho/\rho)_\perp$ at 8,500 oersteds does not represent complete saturation conditions. The horizontal lines in the graph near the origin represent the limits between the values obtained in the region of 600 oersteds in a large number of readings. As

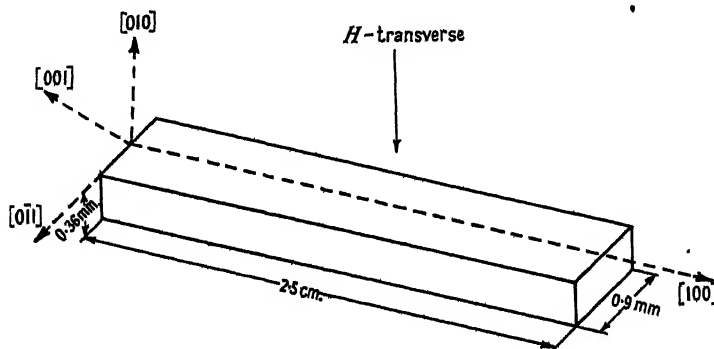


Figure 7. Diagram of single crystal specimen cut along an easy direction of magnetization.

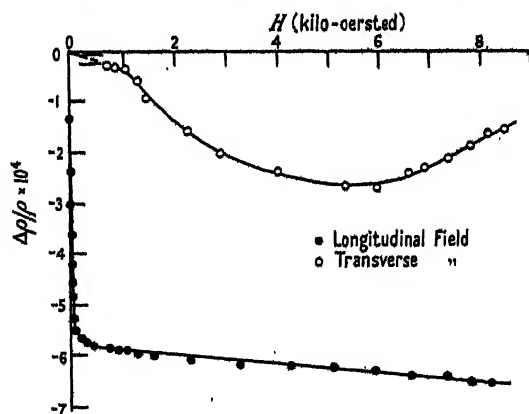


Figure 8. Magneto-resistance curves for specimen of Figure 7.

this specimen had an easy direction of magnetization along its axis we should expect it to saturate in low fields. It will be seen that the change in longitudinal resistance in this specimen reached its final value in a much lower applied field than all the other specimens.

It is interesting to note that the longitudinal coefficient is definitely more negative than the transverse one. Previous workers found, with all materials which they studied, that *under saturation conditions* the quantity $(\Delta\rho/\rho)_\parallel - (\Delta\rho/\rho)_\perp$ was positive; for the present specimen it is negative under these conditions. The other single crystal specimens described above gave positive values of 2 to 5×10^{-4} , whereas the value for the polycrystalline specimen was 1×10^{-4} . If we assume that at room temperature there is no significant variation of magneto-resistance with silicon content from one specimen to another in this work, we must conclude that some crystallites contribute negative saturation values of $(\Delta\rho/\rho)_\parallel - (\Delta\rho/\rho)_\perp$.

§4. SUMMARY AND CONCLUSIONS

- (a) The method described for measuring changes in low resistance can be made sufficiently sensitive to detect changes of 10^{-8} ohm.
- (b) The saturation magneto-resistance of 1.8% silicon-iron fits well into Shirakawa's $(\Delta\rho/\rho)_{||}$ measurements at room temperature for silicon added to iron. The curve has a marked minimum in the region 200–400 oersteds. The value of $B \simeq 10^{-4}$ is unusually small.
- (c) For 1.8% silicon-iron there is no simple relation between $\Delta I/I$ and $(\Delta\rho/\rho)_{||}$, the former ratio varying with field in a manner similar to that for pure iron, the latter in a different manner.
- (d) From single crystal measurements it is concluded that the small $\Delta\rho/\rho$ coefficient is due to a genuine reduction in the contribution of each crystallite, rather than to an averaging-out process.
- (e) A single crystal specimen was prepared for which $(\Delta\rho/\rho)_{||}$ was more negative than $(\Delta\rho/\rho)_{\perp}$. A similar result has so far not been recorded in other materials, possibly due to the lack of data on $(\Delta\rho/\rho)_{\perp}$ for alloys.
- (f) Based on the assumption of a simple magnetization process, as revealed by powder photographs, no simple correlation between $\Delta\rho/\rho$ and α is possible; hence Döring's assumption, $\rho = f(\alpha^2, \alpha^4)$, cannot be applied to silicon-iron without introducing an unjustified approximation.
- (g) In agreement with Kaya's results it was found that in a single crystal the decrease in ρ due to spontaneous magnetization is not necessarily the same for the longitudinal and the transverse case. Considerable variations occur also along different crystal directions. The order of magnitude of this reduction is 50 times smaller than that for nickel in corresponding fields.

ACKNOWLEDGMENTS

The author wishes to thank Professor L. F. Bates, who suggested this problem, for his interest in the work and for many valuable suggestions. The crystals were kindly provided by Mr. G. C. Richer and Mr. A. E. de Barr. Thanks are due to Mr. E. W. Lee for the magnetostriction measurements.

REFERENCES

- BATES, L. F., 1946, *Proc. Phys. Soc.*, **58**, 153.
 BATES, L. F., and NEALE, F. E., 1949, *E.R.A. Report N/T55*; 1950, *Proc. Phys. Soc. A*, **63**, 374.
 BECKER, R., and DÖRING, W., 1939, *Ferromagnetismus* (Berlin: Springer), pp. 311 to 324.
 DÖRING, W., 1938, *Ann. Phys., Lpz.*, **32**, 259.
 DROZHIMA, V. I., and SHUR, V. S., 1947, *C. R. Acad. Sci. U.R.S.S.*, **578**, 1017.
 ENGLERT, E., 1932, *Ann. Phys., Lpz.*, **14**, 589.
 GERLACH, W., 1931, *Ann. Phys., Lpz.*, **8**, 649; 1932a, *Ibid.*, **12**, 849; 1932b, *Phys. Z.*, **33**, 953.
 KAYA, S., 1927, *Sci. Rep. Tôhoku Univ.*, **17**, 1027.
 KORNETZKI, M., 1943, *Ann. Phys., Lpz.*, **43**, 203.
 MCKEEHAN, 1940, *Phys. Rev.*, **36**, 948.
 MOTT, N. F., 1936, *Proc. Roy. Soc. A*, **156**, 368.
 NÉEL, L., 1944, *J. Phys. Radium*, **5**, 241.
 POTTER, H. H., 1931, *Proc. Roy. Soc. A*, **132**, 560.
 SHIRAKAWA, Y., 1939, *Sci. Rep. Tôhoku Univ.*, **27**, 255; 1940, *Ibid.*, **29**, 132, 152.
 STIERSTADT, O., 1930, *Z. Phys.*, **65**, 575; 1930, *Phys. Z.*, **31**, 561.
 WEBSTER, L. W., 1926, *Proc. Roy. Soc. A*, **113**, 196; 1927, *Ibid.*, **114**, 611.
 WILLIAMS, H. J., BOZORTH, R. M., and SHOCKLEY, W., 1949, *Phys. Rev.*, **75**, 155.

Investigations on the Reversible Susceptibility of Ferromagnetics

BY R. S. TEBBLE AND W. D. CORNER *

Physics Department, University of Leeds

Communicated by R. Whiddington; MS. received 7th June 1950

ABSTRACT. A mutual inductometer bridge method has been developed to measure the reversible susceptibility κ_r of a ferromagnetic by the application of small alternating fields to specimens in the form of long wires. The corrections which must be applied because of the finite amplitude of the alternating field and for eddy current effects are investigated and an overall accuracy within about 1% is attained. Typical results are given for Swedish iron which show a marked increase in susceptibility after decarburizing and annealing, and this is discussed in relation to the ideas of Néel. By the evaluation of $\int \kappa_r dH$ over a suitable range of field, the minimum contribution of reversible processes to the total change in magnetization is estimated and found to vary from 10 to 20% for the specimens of iron and nickel examined. Finally an investigation has been made into the statement of Gans that reversible susceptibility is a unique function of intensity of magnetization and independent of the magnetic history of the specimen; the statement is shown to be true only for specimens which have been brought into a given state of magnetization round a hysteresis cycle, and then only for the regions of higher magnetization.

§ 1. INTRODUCTION

THE change with field of the bulk magnetization of a ferromagnetic is due to diverse elementary changes which fall into two main groups. The first group includes continuous movements of domain boundaries through successive equilibrium positions, rotations against the crystal anisotropy forces, and changes in the magnitude of the spontaneous magnetization, all essentially of a reversible character. The second group includes all irreversible boundary movements, whether the movement of an extended boundary across a domain to a position of lower potential energy, or the straightening of a boundary after it has passed some inhomogeneity in its otherwise reversible movement across a domain. These changes are irreversible in a thermodynamic sense too, for they are always accompanied by an irreversible development of heat through the intermediary of the eddy currents induced in the body of the ferromagnetic. The reversible changes approach closely the conditions for strict reversibility in the thermodynamic sense if they are carried out sufficiently slowly, so that eddy current effects are reduced effectively to zero. For irreversible changes, in contrast, once a movement has been initiated it cannot be controlled, and a change occurring during a small increase of field is not reversed by a subsequent corresponding decrease.

Measurements on these two distinctive types of change should yield information which can contribute to a more precise knowledge of the mechanisms by which changes in magnetization occur. Some part, at least, of the irreversible change is associated with the Barkhausen effect, of which a number of experimental studies have been made. Reference to these, and an account of direct measurements of the irreversible changes, are given in papers by Bush and Tebble (1948) and Tebble, Skidmore and Corner (1950); it is shown that, in general, there is a substantial fraction of the whole of the change in magnetization which cannot be ascribed to the Barkhausen processes.

* Now at Durham Colleges in the University of Durham.

The present investigations were begun partly to supplement those on discontinuous processes, and partly to furnish information which might be of use in the interpretation of the magneto-caloric data of Bates and others (see Bates 1949 and Stoner and Rhodes 1949). Section 2 of the present paper contains a discussion of some of the previous measurements of reversible susceptibility. In §3 the methods available for measurement of this quantity are discussed and a full description of one of the methods used in these investigations is given; the effects of eddy currents and demagnetizing factors on the methods are considered in §§4 and 5. The results obtained by the use of this method are discussed in §6 in connection with present ideas on domain structure, with the results of work on the Barkhausen effect and with the implications of Gans's law.

A second method has been developed which is of particular use in investigating variations of susceptibility with temperature. Details of this method and the results obtained in its use, which are of immediate interest in connection with the magneto-caloric effect, will be reported in a separate paper.

§2. DEFINITIONS AND OUTLINE OF PREVIOUS WORK

Definitions.

If the magnetization in a specimen changes by an amount ΔI when the field is altered by ΔH (this being the field corrected for any demagnetization effects due to the form of the body considered), then a susceptibility may be defined[†] as $\Delta I/\Delta H$. Reversible susceptibility, κ_r , at a point specified by (I, H) on an initial or hysteresis curve has usually been defined by

$$\kappa_r = \lim_{\Delta H \rightarrow 0} \Delta I/\Delta H \quad \dots\dots (1)$$

where the field increment ΔH is in the *opposite* direction to the last increment of the field bringing the specimen to the point (I, H) . (The same equation holds for the differential susceptibility, κ_d , but with ΔH in the *same* direction.) An accurate measurement of κ_r directly on the basis of this definition is obviously, in general, difficult, and in most of the more recent work it has been tacitly assumed that the same reversible susceptibility is measured by the use of alternating fields. The corresponding definition is expressed by the same equation (1), but ΔH is now the amplitude of an alternating field superposed on the steady field, and ΔI the corresponding amplitude of the variation in I . It is this 'alternating-field' definition which has been adopted in the present work. In so far as ΔI varies linearly with ΔH over a not too restricted range of ΔH , the value of κ_r obtainable experimentally is unambiguous and accurately measurable. There is no evidence that it differs from the value obtainable in principle from 'direct-field' measurements, but it is perhaps more certainly a reversible susceptibility. It should be noted, however, that non-linearity does not *necessarily* imply irreversibility, and the values obtained, though consistent with a commonly, if tacitly, accepted definition, should strictly be regarded as a lower limit to the reversible susceptibility, in the sense that the integral of $\kappa_r dH$ over any range of H may not give the entire change of I resulting from reversible changes.

The method chosen for the determination of κ_r must clearly be such that accurate measurements may be made of the changes of I due to 'very small' changes of field. Precisely how small the change must be to satisfy the reversibility condition is a matter for experimental investigation in the light of the formal

relations (e.g. range of linearity of the variations) on the one hand and, on the other, of general knowledge of the character of the elementary processes occurring.

Usually the reversible susceptibility is considerably less than either the 'technical susceptibility' I/H or the differential susceptibility dI/dH . At the beginning of the initial magnetization curve, and at high values of H , however, the reversible and differential susceptibilities are very nearly equal, reversible changes in these regions being dominant.

Previous work.

Ewing (1900) recognized the existence of these different susceptibilities, but the first comprehensive measurements of reversible susceptibility were made by Gans (1911). The results were summarized in the statement (sometimes referred to as Gans's law) that "the reversible susceptibility is a unique function of the magnetization and is independent of the magnetic history of a specimen". Further, a relationship, involving the Langevin function, between the reversible susceptibility at a given magnetization and that of the substance in the demagnetized state was given in terms of a parameter x in the form

$$\left. \begin{aligned} \kappa_r/\kappa_i &= 3 L'(x), \\ I/I_s &= L(x), \\ L(x) &= \coth(x) - (1/x), \end{aligned} \right\} \dots\dots(2)$$

where κ_r and I are the values of reversible susceptibility and magnetization under the conditions of measurement, κ_i is the initial reversible susceptibility, and I_s the saturation magnetization.

The equations (2), which give a monotonic decrease of κ_r/κ_i from unity at $I=0$ to zero at $I=I_s$, were not derived theoretically, but the variation indicated was in reasonably good agreement with a considerable number of Gans's results. A tentative theoretical derivation of the equations was later given by Debye (1925), but the basic assumptions which it appeared necessary to make would not now be regarded as physically valid. More recently, some theoretical support for the relation as an approximation has been provided by Brown (1938, 1939) on the basis of a consideration of reversible boundary movements.

Gans's experimental method was to apply small changes of magnetic field to a specimen and to measure the corresponding change in induction ballistically. The specimens, soft and hard steel and iron and hard drawn nickel, were in the form of ellipsoids. Deviations from Gans's law were ascribed to inhomogeneities in the materials. No tests seem to have been made to ensure that the field changes were sufficiently small to give the true reversible susceptibility. Measurements were made along the initial magnetization curve and round the hysteresis loop, but the region inside the loop was not fully investigated.

The work of Erhardt (1917) is of interest in that he was the first to use an inductance comparison method for reversible susceptibility measurements. Some of his results are in fair agreement with Gans's empirical relation, but he was more concerned with the variation of susceptibility with frequency. No change could be detected up to a frequency of 1 Mc/s. This is a rather surprising result, since eddy current and other effects are known to be appreciable for specimens of the dimensions given.

The work of Sizoo (1929) showed that some of the effects ascribed by Gans to inhomogeneities were due to the use of too large measuring fields. For a ring

specimen of laminated silicon iron the Gans relation held much more closely for a field change ΔH of the order of 10^{-3} oersted than for ΔH of the order of 10^{-1} oersted.

Samuel (1928) found that there was a unique functional relationship between reversible susceptibility and magnetization for cobalt, but that it differed markedly from the Gans relation. Brown (1938) has shown that this could be accounted for by anisotropy in the domain orientation. In Samuel's measurements the reversible susceptibility was determined at points along the initial magnetization curve and round the hysteresis cycle taken between near saturation values of magnetization, but not at points inside the hysteresis loop.

None of these previous measurements gave results which could be assumed to hold for the specimens used in the Barkhausen effect measurements of Bush and Tebble (1948) and Tebble, Skidmore and Corner (1950). It is essential, if a complete self-consistent description of the magnetic properties of any substance is to be obtained, that so far as is possible the identical specimens and conditions should be used for the different types of measurement necessary, since slight differences in degree of purity and conditions of strain can give large differences in the structure-sensitive magnetic properties.

§ 3. EXPERIMENTAL METHOD

Any measurement of susceptibility consists essentially of a measurement of the change in magnetization brought about by a change of field. There are two broad groups of possible methods for the measurement of reversible susceptibility. The first of these includes the 'direct' methods in which the change of magnetization is measured with a ballistic galvanometer or a magnetometer. These methods are by no means ideal for the present purpose since it is necessary to use instruments of a high sensitivity in order to detect the small changes in intensity produced by very small changes of field (a thousandth of the coercive force or less).

The second group includes methods based on measurements of the inductance of a coil with and without the specimen under test as a core. The inductance of the coil is obviously a function of the susceptibility of the material of the core, and this can be calculated from the measurements. A variation of this method, which has considerable practical advantages, is the measurement of the mutual inductance of two coils carrying the specimen. The mutual inductance may be balanced against a standard in some form of A.C. bridge circuit and so directly determined, or the voltage induced across the secondary coil may be measured. The bridge system has been used in the work described in this paper. This group of methods has the advantage that within limits the sensitivity increases with the frequency of the applied voltage; eddy current effects, however, set an upper limit to the frequency which can advantageously be used.

The full bridge circuit diagram is shown in Figure 1 and a simplified circuit in Figure 2.

The circuit, based on a Campbell mutual inductometer, is similar to one used by Hartshorn (1925) for mutual inductance measurements and has the advantage of giving direct readings.

In Figure 2, M_c and M are the mutual inductances of the inductometer and coil assembly respectively. The conditions of balance are

$$M = M_c \quad \dots\dots (3a)$$

$$r + \sigma_1 + \sigma_2 = 0, \quad \dots\dots (3b)$$

where σ_1 and σ_2 are effective resistances giving rise to induced voltages in the secondary in quadrature with the purely inductive E.M.F.'s; the effect of σ_1 and σ_2 is balanced by the variable resistance r . The value of the mutual inductance M can be obtained directly from a reading on the inductometer.

In Figure 1 the balance is attained by adjusting the resistance circuit R_3 and R_{4a} , R_{4b} until (3b) is satisfied, and then the inductometer until (3a) is satisfied. As σ_1 and σ_2 may be such as to give rise to a voltage leading or lagging with respect to the inductive component, the switch S enables the balancing potential to be reversed. The current passing in the primary circuit, and hence the magnitude of the applied alternating field, can be measured. Using an A.F. amplifier and headphones, the inductometer can be balanced with an accuracy of better than 1 in 1,000.

The coil assembly consists of a 'secondary' coil wound on a glass tube inside which the specimen is placed. This is mounted coaxially inside a long solenoid (the 'primary' coil) which gives a uniform field over the volume enclosed by the secondary coil. A third coil (the 'polarizing' coil) wound over the primary coil

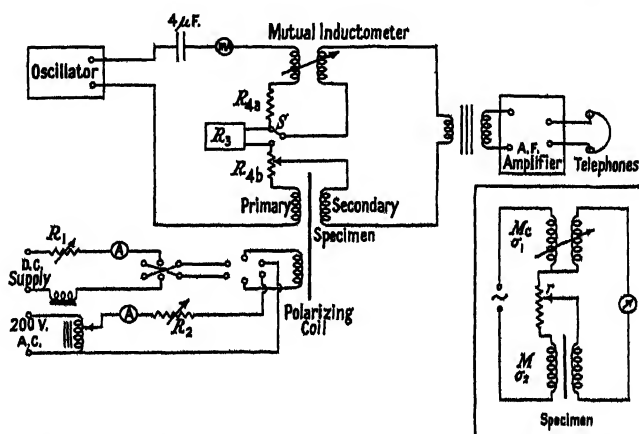


Figure 1. Bridge circuit for measurement of reversible susceptibility.

Figure 2. Simplified bridge circuit.

enables a steady field to be applied to the specimen so as to bring it to any desired point on the magnetization curve; this coil is also used to demagnetize the specimen by means of a slowly reduced alternating current. The use of a choke of high impedance in the supply circuit to the polarizing coil renders negligible the effect of any impedance from this source reflected into the mutual inductance circuit. The assembly is carefully positioned so that the axis of the specimen is at right angles to the earth's magnetic field. The secondary coil is removable and various coils of different diameters and numbers of turns are used to give the required sensitivity for different specimens.

It may be simply shown that the reversible susceptibility with a coil assembly such as the above is given by

$$\kappa_r = \frac{A}{4\pi a} \left(\frac{M_2}{M_1} - 1 \right) \quad \dots (4)$$

where M_2 and M_1 are the values of mutual inductance with and without a specimen present respectively, A is the mean area of cross section of the secondary coil and a that of the specimen.

The accuracy obtainable in the measurement of M_2/M_1 with this apparatus, at least 1 in 1,000 (Rayner and Ford 1950), is much better than can normally be obtained by the alternative method of measuring the voltage induced in the secondary coil. However, the former method is too slow to allow of its use in the susceptibility-temperature measurements, and it may be appropriate to mention here that a rapid automatic method has been devised for this purpose. The voltage induced in the secondary coil is amplified, rectified and displayed as the y deflection of a cathode-ray oscillograph, while the voltage across the polarizing coil provides the x deflection as the specimen is taken through a hysteresis cycle; a (κ_r, H) curve is thus traced on the cathode-ray tube screen and recorded photographically. Fuller details will be given in the later paper.

§ 4. EDDY CURRENT EFFECTS

The assumption has so far been made that the variations in B and H are exactly in phase. This is not generally the case since eddy currents retard the flux change and make the distribution of flux across the specimen non-uniform. The effect is analogous to the skin effect in conductors carrying high-frequency currents, but owing to the high permeability of ferromagnetics it becomes marked at much lower frequencies than does the skin effect in copper conductors; it may be formally represented by treating the permeability as a complex quantity of the form

$$\mu = \mu_0(\alpha + i\beta)$$

with

$$\mu = 4\pi\kappa + 1.$$

It may be shown that the coefficients have the values

$$\alpha = \frac{2}{mr} \left(\frac{\text{ber } mr \text{ bei}' mr - \text{bei } mr \text{ ber}' mr}{\text{ber}^2 mr + \text{bei}^2 mr} \right),$$

$$\beta = \frac{2}{mr} \left(\frac{\text{ber } mr \text{ ber}' mr - \text{bei } mr \text{ bei}' mr}{\text{ber}^2 mr + \text{bei}^2 mr} \right),$$

where $m = (4\pi\omega\mu_0/\rho)^{1/2}$, r is the radius of the specimen, ρ the resistivity, and $\omega/2\pi$ the frequency of the alternating field (Scott 1930).

The voltage induced in the secondary coil is thus a complex quantity, but only the reactive portion of this voltage is balanced by the mutual inductometer and so measured. Thus the true reversible susceptibility is given by the formula

$$\mu_0 = \frac{1}{\alpha} \left[\left(\frac{M_2}{M_1} - 1 \right) \frac{A}{a} + 1 \right].$$

In a method which measures the actual voltage (V_2 and V_1) induced across the secondary it is necessary to take into account both components, and this leads to the expression

$$\frac{V_2 - V_1}{V_1} \frac{A}{a} = [\mu_0^2 \beta^2 + (\mu_0 \alpha - 1)^2]^{1/2}.$$

A close approximation for all actual cases is given by

$$\mu_0 = \frac{1}{(\alpha^2 + \beta^2)^{1/2}} \left(\frac{V_2 - V_1}{V_1} \frac{A}{a} + 1 \right). \quad \dots\dots (6)$$

The functions α and $(\alpha^2 + \beta^2)^{1/2}$ are thus a measure of the eddy current correction factors in each case and are shown in Figure 3, plotted against the parameter mr .

The curves are 'universal' and may be applied to any cylindrical specimen for which the demagnetizing factor is small.

§ 5. EFFECTS DUE TO DEMAGNETIZING FIELD

The effective susceptibility of the specimen is reduced by the effect of the demagnetizing field. The relation between the true reversible susceptibility, κ_r , and the effective value is given by

$$\kappa_r = \kappa_0 \left[\frac{1}{1 - N\mu_0/4\pi} \right],$$

or

$$\kappa_r \simeq \kappa_0 \left[\frac{1}{1 - N\kappa_0} \right],$$

where N is the demagnetizing factor.

To meet the conditions for small eddy current corrections (i.e. small values of mr) the specimen must usually have such a small demagnetizing factor that the effect on the reversible susceptibility measurement is negligible. In the original

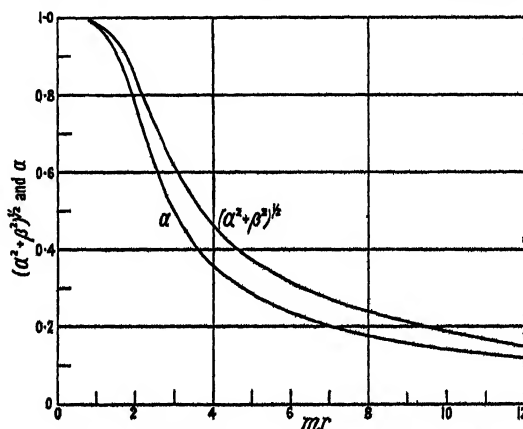


Figure 3. Eddy current correction curves.

design of the apparatus this was considered, and it was constructed to take wire specimens of great (length/diameter) ratio, so that the demagnetizing factor is very small.

§ 6. EXPERIMENTAL RESULTS

(a) Great care was always taken to ensure that the alternating field was small enough for the quantity measured to be a truly reversible susceptibility. The points at which there is the greatest danger of this field being too high are at the foot of the initial magnetization curve and at values of the polarizing field of the order of the coercive force, i.e. where values of κ_r are greatest. Under near saturation conditions it is possible to use large alternating fields with little change in the measured susceptibility. This would be expected if reversible processes predominate in that region. If, therefore, the magnitude of the alternating fields is sufficiently low to satisfy the conditions of reversibility at the critical points mentioned it will be satisfactory for all other parts of the hysteresis loop.

Measurements of the reversible susceptibility were made for various amplitudes of the alternating field at the two critical points and a graph plotted in each case. A typical curve obtained for a demagnetized specimen of hard drawn Swedish iron

is shown in Figure 4. In this case an accuracy to within 1% would be obtained if an alternating field of less than 10^{-2} oersted r.m.s. were employed. In some cases the fields required were lower than could be conveniently used if an accurate balance were to be obtained on the inductometer. In these cases a reading was taken at each of two small values of alternating field amplitude and the value for an

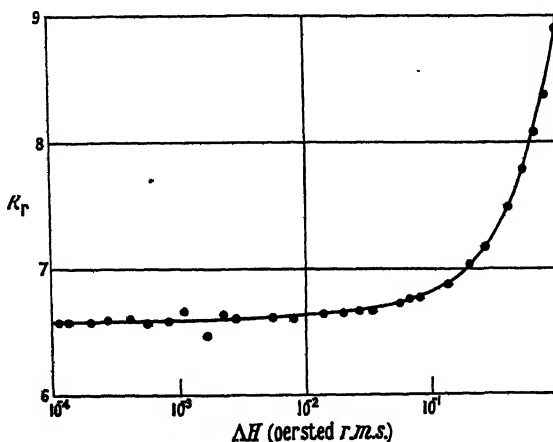


Figure 4. Effect of amplitude of applied alternating field ΔH on measured reversible susceptibility of hard drawn iron.

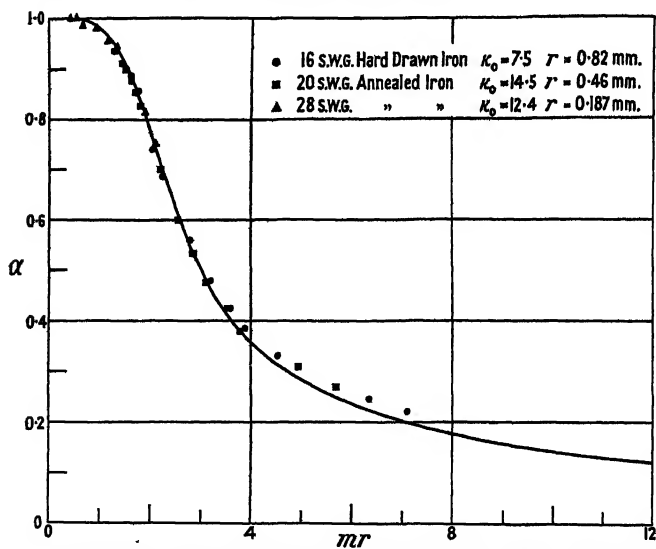


Figure 5. Comparison of calculated and measured eddy current correction factor α .

infinitely small field obtained by extrapolation, assuming the relation between reversible susceptibility and field amplitude to be linear. This is in fact so over a wide range, in agreement with the Rayleigh law. In the measurements an overall accuracy of within $\pm 1\%$ was aimed at.

(b) It was considered desirable to check the eddy current theory experimentally before using it to correct the measured susceptibility for effects due to the frequency of the alternating field. The susceptibility was measured at frequencies up to about 10 kc/s. for three specimens of very different dimensions and susceptibility; the results are shown in Figure 5.

The experimental points lie very close to the theoretical curve for values of $mr < 4$. In the measurements reported here mr has not exceeded 3.5, so that application of the theory is entirely justified.

There is no evidence that the magnetic after-effect discussed by Becker (1939) and Snoek (1947) plays any significant part in the variation of susceptibility under the conditions employed for these measurements. If such an effect were operative it would be expected to cause a further reduction of susceptibility as frequency increased, whereas in fact the susceptibility is rather higher in that region than would be expected on theoretical grounds.

(c) Measurements have been made on a number of materials, but a general indication of the nature of the results is shown in those given for Swedish iron in different stages of annealing. The original specimen, in the form of 26 s.w.g. hard drawn wire, was of composition 99.92% Fe, 0.03% C, 0.05% Mn. It was first decarburized by heating in an electric furnace at a temperature of 950° c. for 48 hours with an atmosphere of moist hydrogen. After cooling it was stretched by 3.25% of its length and then carefully annealed by heating at 850° c. for 72 hours in dry hydrogen so as to induce crystal growth; this has the effect of increasing the grain size by a factor of about 100. The effect of this treatment on the magnetization curves is shown in Figure 6. Measurements of the variation of reversible susceptibility and of magnetization with polarizing field H were made on the specimen at three stages in the treatment: in the initial hard drawn state, after decarburization, and after final annealing to increase grain size.

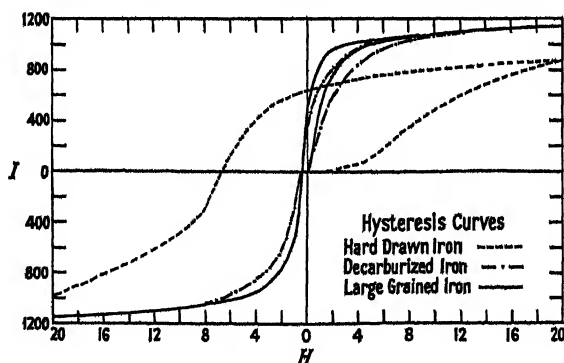


Figure 6. Hysteresis curves for hard drawn iron, decarburized iron and large grained iron.

The results, shown in Figure 7, give the characteristic (κ_r, H) curves. The exact form of the (κ_r, H) relation has been discussed by other workers (see §2), but a detailed consideration of the problem is deferred until the susceptibility-temperature results are presented.

The effect of the decarburizing treatment, which must have included some degree of annealing, is very noticeable and is considerably greater than the effect of the subsequent increase in grain size. This is in accordance with the general ideas, which, as developed quantitatively by Néel (1947), lead to the conclusion that in iron the coercive force, and hence the reversible susceptibility, is affected more by the presence of inclusions of foreign material than by internal strains.

The results for nickel are of a similar character but with values of susceptibility in general somewhat lower; in a well annealed specimen for instance $\kappa_r \sim 10$, with correspondingly lower values for cold-worked specimens. In all the materials

examined the reversible susceptibility was found to be extremely structure-sensitive, and great care was necessary in handling specimens, particularly those which had been subjected to any degree of annealing. In fact the measurement of reversible susceptibility in the region $H \sim H_0$ is a very delicate test as to whether any significant structural change has taken place in a material.

(d) A feature of interest in susceptibility measurements lies in the fact that the area under a (κ_r, H) curve, given by

$$\int \kappa_r dH = \int \frac{dI_R}{dH} dH = \Delta I_R$$

with appropriate limits, is the contribution of the reversible processes to the total change in magnetization over the range of field considered.

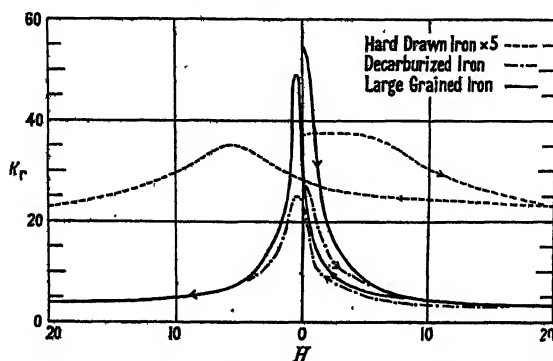


Figure 7. Variation of reversible susceptibility κ_r with polarizing field H .

The Table gives a summary of results on specimens for which information is also available on the contribution of the Barkhausen effect ΔI_B to the total change in magnetization ΔI_T . The sum $\Delta I_R + \Delta I_B$ should give a magnetization change equal to the total ΔI_T .

Contribution by Discontinuous and Reversible Processes (ΔI_B and ΔI_R)
to the Total Change in Magnetization ΔI_T .

Material	Limits of field	ΔI_T	ΔI_B	ΔI_R	$\Delta I_B + \Delta I_R$
Hard drawn iron	± 18.6	2150	1844	180	2024
Large grained iron	± 3.0	2180	1250	110	1360
Annealed nickel	± 14.9	700	326	115	441

It will be seen that only in one case is the sum of ΔI_B and ΔI_R closely equal to ΔI_T . This failure to account for the whole of the change in terms of the measured reversible and irreversible changes has been discussed by Tebble, Skidmore and Corner (1950). The main reason is considered to be the insufficient sensitivity of the apparatus used for the detection of the Barkhausen effect in the large grained iron and annealed nickel specimens; the minimum detectable volumes correspond to $0.87 \times 10^{-10} \text{ cm}^3$ and $3.1 \times 10^{-10} \text{ cm}^3$ respectively. It is thought probable that the effect of the annealing and increase in grain size in the preparation of the large grained iron is to reduce the number of the larger Barkhausen discontinuities and to increase the number of the smaller irreversible processes which lie below the range of the apparatus. There is no indication that the method of estimating the reversible contribution is seriously at fault, though, as mentioned earlier (§ 2),

the total reversible change in I may be somewhat greater than that estimated from the measured reversible susceptibility.

(e) A series of measurements designed to test the validity of the Gans relation was carried out over a wide range of conditions, as shown in Figure 8. This represents the upper half of a series of hysteresis cycles round which a specimen of 26 s.w.g. hard drawn iron was taken, the reversible susceptibilities being measured at suitable points on each curve. Besides the normal type of hysteresis loop, with

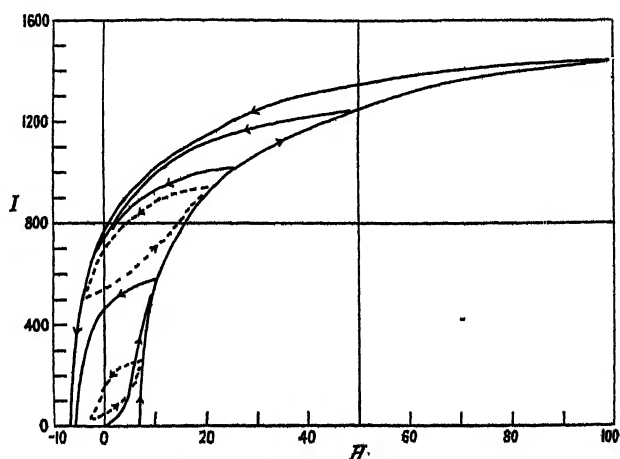


Figure 8. Magnetization curves for hard drawn iron.

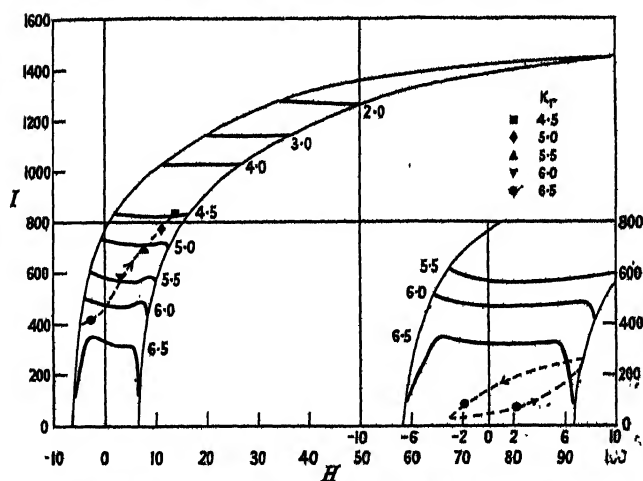


Figure 9. Contour diagrams for hard drawn iron.
Values of susceptibility are shown by figures against contours.

field limits of 100, 50, 25 and 10 oersteds, measurements were also carried out on the two subsidiary loops shown in Figure 8; a further path on the (I, H) diagram was also traversed from $H = -5.2, I = 400$ to $H = +14.1, I = 840$, as shown in Figure 9. Using these results, a contour diagram (Figure 9) was constructed, the contours joining points of equal reversible susceptibility. If Gans's law were strictly true all the contours would be straight lines parallel to the H -axis. This is approximately true for the results of the measurements on the loops for higher values of magnetization, but for the lower values the divergence from linearity is marked.

The results of the measurements on the path $(-5.2, +400)$ to $(+14.1, +840)$ and those on the smaller of the subsidiary loops are shown separately in Figure 9, as the disagreement in these cases is particularly noticeable. On these curves the points at which the susceptibility is of the same magnitude as on a contour are shown using a characteristic mark.

Now the importance of Gans's law lies not so much in the form of the equation (2) *per se* but in the statement that κ_r is a unique function of I . It would appear from these results, and from an examination of the susceptibility curve, that this holds only for levels of magnetization above that at which the initial magnetization curve and the ascending portion of the hysteresis loop become concurrent. However, even this is only true for measurements taken on specimens brought into a given position on the (I, H) plane by proceeding in the normal manner round a hysteresis cycle. A path such as that indicated in Figure 9 will give results at variance with Gans's law even for higher values of magnetization.

The reversible susceptibility is thus not independent of the path by which a specimen is brought into a given state, nor is it a unique function of field or magnetization or of both except in the high magnetization region or where the path followed is of a defined type. This result is perhaps not surprising since approaches to a given state in different ways will involve different types of elementary processes, and the magnetization may be the resultant of quite different microscopic conditions. It might be expected that in regions where a particular type of change is dominant there would be considerable similarity in the internal state of a specimen to produce a given magnetization; this is found for high magnetizations where reversible effects predominate and where Gans's law is followed with reasonable accuracy.

ACKNOWLEDGMENTS

The authors wish to thank Professor R. Whiddington, under whose direction this work has been carried out, for his interest and support and Professor E. C. Stoner for his help in the preparation of the paper. The samples of nickel and iron were kindly provided by the Mond Nickel Company and Messrs. Rigby and Sons Ltd.

REFERENCES

- BATES, L. F., 1949, *J. Phys. Radium*, **10**, 353.
 BECKER, R., 1939, *Ann. Phys., Lpz.*, **36**, 340.
 BROWN, W. F., 1938, *Phys. Rev.*, **54**, 278; 1939, *Ibid.*, **55**, 568.
 BUSH, H. D., and TEBBLE, R. S., 1948, *Proc. Phys. Soc.*, **60**, 370.
 DEBYE, P., 1925, *Handbuch der Radiologie*, **6**, 720 (Leipzig).
 ERHARDT, F., 1917, *Ann. Phys., Lpz.*, **54**, 41.
 EWING, J. A., 1900, *Magnetic Induction in Iron* (London: Electrician Series).
 GANS, R., 1911, *Phys. Z.*, **12**, 1053.
 HARTSHORN, L., 1925, *J. Sci. Instrum.*, **2**, 145.
 NÉEL, L., 1947, *Ann. Univ. Grenoble*, **22**, 299.
 RAYNER, G. H., and FORD, L. H., 1950, *J. Sci. Instrum.*, **27**, 19.
 SAMUEL, M., 1928, *Ann. Phys., Lpz.*, **86**, 798.
 SCOTT, K. L., 1930, *Proc. Inst. Radio Engrs*, **18**, 1750.
 SIZOO, G. J., 1929, *Ann. Phys., Lpz.*, **3**, 270.
 SNOEK, J. L., 1947, *New Developments in Ferromagnetic Materials*. Monographs on Progress of Research in Holland (Amsterdam: Elsevier).
 STONER, E. C., and RHODES, P., 1949, *Phil. Mag.*, ser. 7, **40**, 481.
 TEBBLE, R. S., SKIDMORE, I. C., and CORNER, W. D., 1950, *Proc. Phys. Soc. A*, **63**, 739.

Congruent Space Charge Flow

By G. B. WALKER

Department of Electrical Engineering, Imperial College, London

Communicated by Willis Jackson; MS. received 24th March 1950

ABSTRACT. The paper is mainly concerned with one type of congruent flow, namely irrotational flow, and propositions are established regarding (a) rectilinear motion, (b) motion in which the current density is constant along lines of flow and (c) the representation of lines of flow by the level lines of a harmonic function. From the latter, three corollaries are deduced regarding flow in which the space charge density is constant either throughout the motion, or along lines of flow or along lines of constant action.

Two cases of curvilinear flow, originating from a unipotential cathode, are discussed and are shown to possess important features regarding magnification and transit time.

§ 1. INTRODUCTION

IN a large class of electronic devices electrons move as a continuous fluid in a stationary field in which the magnetic force is negligible. The flow is congruent in the sense that lines of flow form a system of curves filling a portion of space such that in general a single curve passes through any given point, the velocity in the motion being a single valued function of position. The equations governing the motion are

$$\dot{\mathbf{v}} = (e/m) \text{grad } V \quad \dots\dots(1)$$

$$\nabla^2 V = 4\pi\rho \quad \dots\dots(2)$$

$$\text{div } \mathbf{i} = 0 \quad \dots\dots(3)$$

and, since the flow is congruent

$$\dot{\mathbf{v}} = (\mathbf{v} \cdot \nabla) \mathbf{v} = \frac{1}{2} \text{grad } (v^2) - \mathbf{v} \times \text{curl } \mathbf{v} \quad \dots\dots(4)$$

where V is the electric potential, v the velocity, e and ρ the scalar magnitudes of the electron charge and the space charge density respectively, m the mass of an electron and $\mathbf{i} (= \rho \mathbf{v})$ the current density.

There exists no general method by which these equations can be integrated rigorously but it is possible to make some deductions concerning particular solutions and concerning types of motion which have specific physical characteristics.

§ 2. CLASSIFICATION OF THE FLOW

There are two characteristic features of the flow, namely that the circulation of the velocity round any closed circuit moving with the flow is constant (as in hydrodynamics, see for example, Weatherburn 1937) and that along a line of flow the sum of the potential and kinetic energy of an electron is constant (see, for example, Meltzer 1949). In a recent paper Meltzer (1949) has distinguished two fundamental types of congruent flow which he has named normal and abnormal, the former being defined as a flow in which the sum of the potential and kinetic energy is everywhere constant. He has shown, moreover, that a necessary and sufficient condition for 'normal' flow is that $\mathbf{v} \times \text{curl } \mathbf{v}$ is everywhere zero.

There are other ways in which the flow can be classified. For example, in geometrical terms a normal, as distinct from a skew, congruence refers to the case in

which there exists a singly infinite family of surfaces which are cut orthogonally by the lines of flow. A necessary and sufficient condition for this is that $\mathbf{v} \cdot \text{curl } \mathbf{v}$ vanishes, as may easily be deduced from geometrical considerations (Synge 1937). Alternatively, in dynamical terms one may distinguish between irrotational and rotational (or vortex) motion, the condition for the former being $\text{curl } \mathbf{v} = 0$. Clearly, these three ways of classification differ only in their terms of reference and there is no question of one being more fundamental than the other, since the three are practically identical, the last being the most stringent since it is contained in the other two.

The limitations on the practical realization of congruent flow are very considerable since in all known processes of emission electrons have a statistical velocity distribution rendering such flow impossible unless it can be assumed that the initial velocity is negligible. It is sometimes assumed that electrons are ejected normal to the cathode with a constant velocity (depending on the cathode temperature). This may partially be justified in dealing with rectilinear motion but has no meaning in any other case.

Only two types of zero velocity emitter can be envisaged, an equipotential surface such as a normal indirectly heated thermionic cathode, or a varying potential surface such as a varying potential directly heated cathode. For the former the flow is 'normal' in Meltzer's sense since at the emitter the sum of the potential and kinetic energy is constant. Furthermore, if $\text{curl } \mathbf{v}$ exists it must be directed along a line of flow and hence vortex tubes coincide with tubes of flow. The vorticity is measured by the circulation of the velocity round any closed circuit lying on a vortex tube and circumscribing it once. Such a circuit can be found which also lies on the emitting surface and consequently the vorticity is zero. All three normality criteria are thus satisfied. Likewise if the emitter is not an equipotential, $\text{curl } \mathbf{v}$ has a value (becoming infinite at the emitter) and the flow is 'abnormal', skew and vortex.

§ 3. IRROTATIONAL FLOW

In irrotational flow the velocity can always be expressed as the gradient of a scalar point function, namely the action function W ,

$$\text{i.e.} \quad \mathbf{v} = \text{grad } W. \quad \dots\dots(5)$$

By combining (1), (4) and (5) it follows that

$$\text{grad } (\text{grad } W)^2 = (2e/m) \text{ grad } V$$

$$\text{hence} \quad (\text{grad } W)^2 = (2e/m) V + C \quad \dots\dots(6)$$

where C is a constant which may be put equal to zero without loss of generality, the consequence merely being that speeds are zero when the potential is zero.

In the case of two-dimensional motion a necessary and sufficient condition that (3) be satisfied is that a stream function ψ exists such that

$$\left. \begin{aligned} i_x = \rho \frac{\partial W}{\partial x} &= \frac{\partial \psi}{\partial y} \\ i_y = \rho \frac{\partial W}{\partial y} &= -\frac{\partial \psi}{\partial x} \end{aligned} \right\} \quad \dots\dots(7)$$

where i_x and i_y are the rectangular components of the vector current density. Each line $\psi = \text{const.}$ is orthogonal to any line $W = \text{const.}$ and so the functions ψ

and W may be used as a reference system of orthogonal curvilinear coordinates. By referring to the elementary cell shown in Figure 1

$$dS_1 = h_1 dW; \quad dS_2 = h_2 d\psi \quad \dots\dots(8)$$

where (Stratton 1940)

$$h_1^2 = \left(\frac{\partial x}{\partial W}\right)^2 + \left(\frac{\partial y}{\partial W}\right)^2; \quad h_2^2 = \left(\frac{\partial x}{\partial \psi}\right)^2 + \left(\frac{\partial y}{\partial \psi}\right)^2.$$

From equations (7) it follows that

$$i^2 = \rho^2 \left[\left(\frac{\partial W}{\partial x}\right)^2 + \left(\frac{\partial W}{\partial y}\right)^2 \right] = \left(\frac{\partial \psi}{\partial x}\right)^2 + \left(\frac{\partial \psi}{\partial y}\right)^2 \quad \dots\dots(9)$$

and hence, by expressing $\partial x/\partial W$, $\partial y/\partial W$, etc. in terms of $\partial W/\partial x$, $\partial W/\partial y$, etc. it follows that

$$h_1^2 = \frac{1}{\rho^2}; \quad h_2^2 = \frac{1}{i^2}. \quad \dots\dots(10)$$

The operator ∇^2 can then be written

$$\begin{aligned} \nabla^2 &= \frac{1}{h_1 h_2} \left[\frac{\partial}{\partial W} \frac{h_2}{h_1} \frac{\partial}{\partial W} + \frac{\partial}{\partial \psi} \frac{h_1}{h_2} \frac{\partial}{\partial \psi} \right] \\ &= \frac{i^2}{\rho} \left[\frac{\partial}{\partial W} \frac{1}{\rho} \frac{\partial}{\partial W} + \frac{\partial}{\partial \psi} \rho \frac{\partial}{\partial \psi} \right] \quad \dots\dots(11) \end{aligned}$$

and the well-known relationship between force and acceleration components perpendicular to a trajectory is

$$i \frac{\partial V}{\partial \psi} = - \frac{2V}{R_\psi} \quad \dots\dots(12)$$

where R_ψ is the radius of curvature of the trajectory.

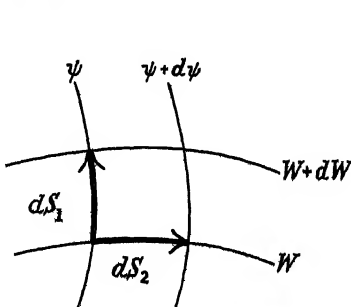


Figure 1. Elementary cell defined by the orthogonal curvilinear coordinates (W , ψ).

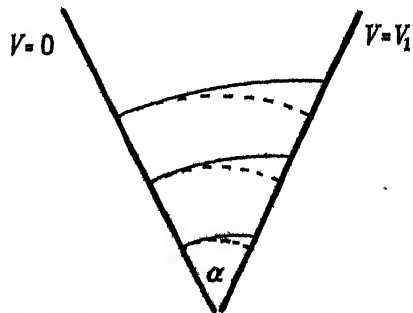


Figure 2. Electron trajectories between two inclined planes. Lines of force are arcs of circles shown by the broken lines.

Three general deductions regarding two-dimensional motion will now be made. For clarity they are stated as propositions.

(i) Proposition on Rectilinear Motion

There are only two possible types of rectilinear motion in two dimensions, namely (a) that in which stream lines are parallel, and (b) that in which stream lines radiate from a point.

For lines of flow to be of zero curvature they must coincide with lines of force and hence (by (12)) V and v are functions of W only. Equipotential lines coincide with lines of constant action and so have the same radius of curvature R at any given point. These lines form a system of parallel curves and from (8) and (10) it follows that

$$\frac{\partial R}{\partial W} = \frac{1}{v}. \quad \dots\dots (13)$$

By means of (11)

$$\nabla^2 W = \frac{i^2}{\rho} \frac{\partial}{\partial W} \frac{1}{\rho} = -\frac{v^2}{\rho} \frac{\partial \rho}{\partial W}. \quad \dots\dots (14)$$

The operator ∇^2 can also be written

$$\nabla^2 = \frac{\partial^2}{\partial n^2} + \frac{1}{R} \frac{\partial}{\partial n} \quad \dots\dots (15)$$

where $\partial/\partial n$ indicates differentiation in a direction normal to a level line of the function operated on and R is the radius of curvature of the level line. Thus

$$\nabla^2 W = \frac{\partial^2 W}{\partial n^2} + \frac{1}{R} \frac{\partial W}{\partial n} = v \left(\frac{dv}{dW} + \frac{1}{R} \right). \quad \dots\dots (16)$$

By combining (14) and (16) it follows that

$$\rho \left(\frac{dv}{dW} + \frac{1}{R} \right) + v \frac{\partial \rho}{\partial W} = 0. \quad \dots\dots (17)$$

This equation must be satisfied identically. In addition* Poisson's equation (2) must also be satisfied, which by means of (15) can be written

$$4\pi\rho = \nabla^2 V = v \left[\frac{d}{dW} \left(v \frac{dV}{dW} \right) + \frac{1}{R} \frac{dV}{dW} \right]. \quad \dots\dots (18)$$

It is now possible to eliminate ρ between (17) and (18), and replacing $\partial R/\partial W$ by $1/v$, cf. (13), there results the identity

$$\left(2 \frac{dv}{dW} \frac{d}{dW} + v \frac{d^2}{dW^2} \right) \left(v \frac{dV}{dW} \right) + \frac{1}{R} \left[\frac{dv}{dW} \frac{dV}{dW} + 2 \frac{d}{dW} \left(v \frac{dV}{dW} \right) \right] \equiv 0. \quad \dots\dots (19)$$

Clearly, therefore, R is a function of W only. Action lines are thus lines of constant curvature and so can only be a family of concentric circles, or in the limiting case, a family of parallel straight lines.

For simplicity this proposition has been stated and proved for two-dimensional motion. In the case of three-dimensional motion action surfaces coincide with equipotential surfaces and form a system of parallel surfaces. By a similar argument to the foregoing it can be shown, by investigating the principal radii of curvature of these surfaces, that there are only three types of rectilinear motion, namely lines of flow (a) parallel, (b) radiating from a point, and (c) radiating normally from an axis.

(ii) *Proposition on Motion with Constant Current Density*

The only motions in which the current density is constant along a line of flow are (a) stream lines which are parallel straight lines, and (b) stream lines which are concentric circles.

The motion is characterized by the fact that $|\mathbf{i}|$ is a function of ψ only. Clearly therefore, since $h_2 = 1/|\mathbf{i}|$ stream lines must form a family of parallel curves. Thus

$$\frac{\partial R_\psi}{\partial \psi} = h_2 = \frac{1}{|\mathbf{i}|} \quad \dots\dots(20)$$

and hence on integration the most general form of R_ψ is

$$R_\psi = \int \frac{1}{|\mathbf{i}|} d\psi + a \quad \dots\dots(21)$$

where a is a function of W only.

If (21) is substituted in (12) there results

$$\frac{1}{2} \frac{\partial \log V}{\partial \psi} = - \frac{1/|\mathbf{i}|}{\int (1/|\mathbf{i}|) d\psi + a} \quad \dots\dots(22)$$

and hence by integration and rearrangement

$$V^{1/2} = \frac{b}{c + a} \quad \dots\dots(23)$$

where b is a function of W only and $c = \int (1/|\mathbf{i}|) d\psi$ is a function of ψ only. From (6), (9) and (23) it follows that the most general form of ρ is

$$\rho = de + fg \quad \dots\dots(24)$$

where d and f are functions of W only, and e and g are functions of ψ only.

If Poisson's equation (2) is now set down with the operator ∇^2 in the form (11) it will be seen to consist of a number of terms each of which is the product of a function of ψ only and a function of W only. It is easy to show that this equation can be satisfied identically only under the alternatives (a) that all the functions of W vanish or are constants, or (b) that all the functions of ψ vanish or are constants.

Under the first alternative it follows from (21) that R_ψ is a function of ψ only, but the ψ -curves are a family of parallel curves and hence must consist of a family of concentric circles. A complete solution can easily be obtained but the case is of little practical importance. Under the second alternative the current density is everywhere constant, and from (21) R_ψ is a function of W only. Since the ψ -curves are parallel curves they must in fact be a family of straight lines.

(iii) Proposition on Harmonic Functions

In a two-dimensional electron stream lines of flow cannot be represented by the level lines of a harmonic function except in trivial cases to be named.

Let w be a harmonic function whose level lines represent lines of flow, and let u be the conjugate of w . Clearly ψ is a function of w only, and W is a function of u only. By means of equations (7) it follows that

$$\rho = \frac{\psi'}{W'} \quad \dots\dots(25)$$

and from (6)

$$V = \frac{m}{2e} [W']^2 \left[\left(\frac{\partial u}{\partial x} \right)^2 + \left(\frac{\partial u}{\partial y} \right)^2 \right]. \quad \dots\dots(26)$$

Since u is harmonic it can be expressed as the sum of a complex analytic function and its conjugate, i.e.

$$u = f + f^* \quad \dots\dots(27)$$

where f is a function of $x + iy$ and f^* is the same function of $x - iy$. Likewise

$$w = -i(f - f^*). \quad \dots\dots(28)$$

Let
$$z = (m/2e)(W')^2 \quad \dots\dots(29)$$

then equation (26) can be written

$$V = 4z f' f'^* \quad \dots\dots(30)$$

and hence

$$\nabla^2 V = 16[(f' f'^*)^2 z'' + \{(f')^2 f''^* + f''(f'^*)^2\} z' + f'' f''^* z]. \quad \dots\dots(31)$$

We must now consider the conditions under which Poisson's equation (2) can be satisfied when $\nabla^2 V$ is given by (31) and ρ by (25). It is convenient to discuss three cases separately, corresponding to the following assumptions about ρ , namely (i) that ρ is zero, (ii) that ρ is a function of u only, i.e. that ψ' is a constant, and (iii) that neither ψ' nor W' are constants.

Case 1. $\rho = 0$.

Since the current density must also be zero it follows that $\text{div } \mathbf{i} = 0$. Poisson's equation reduces to the Laplace equation $\nabla^2 V = 0$ and hence from (31) the condition to be satisfied is

$$(f' f'^*)^2 \left[z'' + \left\{ \frac{f''}{(f')^2} + \frac{f''^*}{(f'^*)^2} \right\} z' + \frac{f''}{(f')^2} \cdot \frac{f''^*}{(f'^*)^2} z \right] \equiv 0. \quad \dots\dots(32)$$

Now, if $f' f'^*$ is zero, V is also zero by (30) and hence motion is impossible. We must therefore examine the conditions under which the expression contained in the square brackets vanishes.

If z is zero, motion is impossible, and if z is a constant there can be no acceleration in the motion, a trivial case. Let us assume therefore that both z and z' exist and proceed to examine their coefficients. The identity can clearly be established if either f'' is zero, or f''/f'^2 is a constant and it can easily be verified that the former refers to a motion in which the acceleration is constant and in a fixed direction, whereas the latter refers to a motion along lines radiating from a point (in fact the only possible cases of rectilinear motion).

If, however, the coefficients of z and z' do not vanish and are not constants we must examine the possibility that they are both functions of u only. Now, the coefficient of z' is real and is a harmonic function (since it is the sum of two complex conjugates) and hence it can only have the form au where a is a constant. This follows from the fact that the only function of a harmonic function which is itself a harmonic function is a constant multiplied by the original harmonic function. But if

$$\frac{f''}{(f')^2} + \frac{f''^*}{(f'^*)^2} = au$$

the coefficient of z , namely

$$\frac{f''}{(f')^2} \frac{f''^*}{(f'^*)^2} = \frac{1}{4} a^2 (u^2 + w^2)$$

is not a function of u only and hence the identity cannot be established.

Case 2. ρ a function of u only.

From (25) it follows that

$$\rho = \frac{b}{W'} = \frac{4c}{\pi z^{1/2}}$$

where b and c are constants, and hence combining equations (6), (2) and (31) the following identity must hold

$$(f'f'^*)^2 \left[z'' + \left\{ \frac{f''}{(f')^2} + \frac{f''^*}{(f'^*)^2} \right\} z' + \frac{f''}{(f')^2} \frac{f''^*}{(f'^*)^2} z \right] \equiv - \frac{c}{z^{1/2}} \dots\dots (33)$$

This identity may be discussed in the same way as the previous one. For example :

(i) If $z=0$ motion is impossible.

(ii) If z is a constant, f'' and f''^* must also be constants and it is easy to show that in the motion lines of flow are hyperbolae. It is important to note that in this case the action function is a harmonic function and the space charge density is constant. This case will be referred to later.

(iii) If z and z' exist then either $f''=0$, in which case lines of flow are parallel, or f''/f'^2 is a constant, in which case lines of flow are straight lines radiating from a point.

By carrying out a similar argument as in case 1 it can be shown that there are no other possible motions.

Case 3. ψ' and W' exist and are not constants

The identity to be established differs from (33) only in that c is now a function of w only. By multiplying both sides by $z^{1/2}$ it is clear that a necessary condition is that

$$z^{1/2}(f'f'^*)^2 \left[z'' + \left\{ \frac{f''}{(f')^2} + \frac{f''^*}{(f'^*)^2} \right\} z' + \frac{f''}{(f')^2} \frac{f''^*}{(f'^*)^2} z \right]$$

should be function of w only. That this is impossible can easily be shown by similar arguments to those set out in the previous cases.

To summarize, it can be stated that stream lines (and hence lines of constant action) can be represented by the plot of the level lines of harmonic function in only two types of motion, namely rectilinear motion and motion along hyperbolic paths. Three important corollaries can also be stated, viz.

Corollary 1. The only possible motion in which the space charge density is constant throughout is hyperbolic motion.

A necessary condition to be satisfied is

$$\text{div } \mathbf{i} = 0 \dots\dots (3)$$

In irrotational flow the components of the vector current density \mathbf{i} are $\rho \frac{\partial W}{\partial x}$ and $\rho \frac{\partial W}{\partial y}$ and hence (3) becomes

$$\rho \nabla^2 W + \text{grad } \rho \cdot \text{grad } W = 0. \dots\dots (34)$$

But if ρ is a constant, $\text{grad } \rho$ is zero and hence $\nabla^2 W = 0$. The motion is thus according to Case 2(ii) above.

Corollary 2. The only motion in which the space charge density is constant along a line of flow is hyperbolic motion.

This follows at once from Corollary 1. The restriction now imposed is that $\text{grad } \rho \text{ grad } W = 0$ and hence by (34) $\nabla^2 W = 0$.

Corollary 3. The only motions in which the space charge density is constant along lines of constant action are rectilinear and hyperbolic motions.

From equations (7) we have

$$\frac{\partial W}{\partial x} = \frac{1}{\rho} \frac{\partial \psi}{\partial y}; \quad \frac{\partial W}{\partial y} = -\frac{1}{\rho} \frac{\partial \psi}{\partial x}. \quad \dots\dots(35)$$

By differentiating the first with respect to y , the second with respect to x and subtracting there results

$$\frac{1}{\rho} \nabla^2 \psi - \frac{1}{\rho^2} \text{grad } \rho \text{ grad } \psi = 0.$$

This is a necessary condition and in the present case there is the further restriction that $\text{grad } \rho \text{ grad } \psi = 0$. It follows therefore that $\nabla^2 \psi = 0$; thus ψ is a harmonic function and hence ψ' must be a constant. The possible motions are thus described under Case 2 above.

§ 4. PARTICULAR SOLUTIONS

If attention is confined as in the foregoing to irrotational flow, the method suggested by Meltzer (1949) for finding particular solutions reduces to the choice for W , the action function, of an arbitrary polynomial or a set of simple known functions. It is unlikely that any new information can be obtained in this way since it is not enough merely to satisfy the field-dynamical equations. Thus, the case of hyperbolic motion given by Meltzer, which has been shown to be the only non-trivial case in which the action function is a harmonic function, namely $W = xy$, is not realizable since the conditions to be satisfied at a cathode are not fulfilled. It can be shown, also, that no real solutions can be obtained by expressing W as a polynomial with a finite number of terms of the type $x^n y^m$ where n and m are integers (since when the velocity is zero the space charge density is finite). It would appear, therefore, that rigorous solutions are known only for the three types of rectilinear motion. There are, however, two cases of practical interest in which the motion is along curved paths and these will now be discussed.

§ 5. MOTIONS BETWEEN TWO INCLINED PLANAR ELECTRODES

The motion of electrons in a region bounded by two plane conductors inclined at an angle possesses some interesting features which do not appear to be known and which may have important applications. Consider first the case of negligible space charge and let it be assumed that the electrons depart from one plane with no initial velocity. With polar coordinates, as shown in Figure 2, the potential in the field can be written $V = V_1 \theta / \alpha$ and equation (6) takes the form

$$\left(\frac{\partial W}{\partial r} \right)^2 + \left(\frac{1}{r} \frac{\partial W}{\partial \theta} \right)^2 = K\theta \quad \dots\dots(36)$$

where K is a constant whose value is $2eV_1/m\alpha$. Since the squared terms must be independent of r we can seek a solution in the form $W = ry(\theta)$. Equation (36) then reduces to

$$y^2 + y'^2 = K\theta \quad \dots\dots(37)$$

which is solved by

$$y = \sum_{n=1}^{\infty} a_n \theta^{(4n-1)/2}$$

The constants a_n can readily be evaluated, for example

$$a_1 = \frac{2}{3} K^{1/2}, \quad a_2 = -\frac{2}{21} a_1, \quad a_3 = \frac{10}{2079} a_1, \text{ etc.}$$

The radial and transverse velocity components are given by

$$\dot{r} = \frac{\partial W}{\partial r} = y; \quad r\dot{\theta} = \frac{1}{r} \frac{\partial W}{\partial \theta} = y'$$

Hence

$$\frac{dr}{r} = \frac{y}{y'} d\theta$$

and by integration

$$r = r_0 \exp \phi(\theta) \quad \dots\dots(38)$$

where

$$\phi(\theta) = \int_0^\theta \frac{y}{y'} d\theta.$$

A series expansion for $\phi(\theta)$ can easily be obtained, namely

$$\phi(\theta) = \sum_{n=1}^{\infty} b_n \theta^{2n}$$

where

$$b_1 = \frac{1}{3}, \quad b_2 = \frac{4}{189}, \quad b_3 = \frac{32}{18711}, \text{ etc.}$$

Clearly the motion is such that the radii of the end points of the trajectories bear the same ratio to each other as do the starting points and so a perfect linear magnification is obtained. The time of transit of the electrons also has an interesting property. To examine this consider the relationship between transverse acceleration and force in a trajectory, namely

$$m \frac{1}{r} \frac{d}{dt} (r^2 \dot{\theta}) = \frac{eV_1}{r\alpha}.$$

On removing $1/r$ from both sides and integrating with respect to time we obtain $t = 2r^2 \dot{\theta} / K$ where t is measured from the instant the electron leaves the first plane. Noting that $r\dot{\theta} = y'$, a function of θ only, it follows that

$$t = r \left(\frac{2}{K} y' \right). \quad \dots\dots(39)$$

Hence the transit time is proportional to the final (or starting) radius.

The above-mentioned properties of the motion both derive from the functional form of the action function. It is interesting to see, therefore, if by writing $W = rg(\theta)$ a possible motion can be defined in which the space charge is not negligible. By substitution in (6)

$$g^2 + g'^2 = \frac{2e}{m} V \quad \dots\dots(40)$$

and hence V is a function of θ only.

By expressing the operator ∇^2 in terms of polar coordinates and applying Poisson's equation (2) it follows that

$$\rho = \frac{V''}{4\pi r^2}. \quad \dots\dots(41)$$

The radial and transverse components of the vector current density are $i_r = V''g/4\pi r^2$ and $i_\theta = V''g'/4\pi r^2$.

$$\begin{aligned}\text{Also} \quad \text{div } \mathbf{i} &= \frac{\partial i_r}{\partial r} + \frac{1}{r} i_r + \frac{1}{r} \frac{\partial i_\theta}{\partial \theta} \\ &= \frac{1}{4\pi r^3} \left[-2V''g + V''g + \frac{d}{d\theta}(V''g') \right].\end{aligned}$$

Therefore, $\text{div } \mathbf{i}$ is zero if

$$\frac{d}{d\theta}(V''g') - V''g = 0. \quad \dots\dots(42)$$

It is now possible to eliminate V between equations (40) and (42) and obtain an ordinary differential equation in g . Any solution of this equation defines a possible motion. By writing

$$g = \sum_{n=1}^{\infty} \alpha_n \theta^{(8n-1)/3}$$

it is easy to show that the constants α_n can be determined such that (40) and (42) are satisfied. Corresponding to this value of g it follows from (40) that V can be expanded in a series of the form

$$V = \sum_{n=1}^{\infty} \beta_n \theta^{(8n-2)/3}.$$

Clearly, therefore, the potential gradient at the cathode (where $\theta=0$) is zero and hence this solution gives the space-charge-saturated case. Proceeding as before it can be shown that the equation of a trajectory is now

$$r = r_0 \exp \left(\frac{3}{10} \theta^2 + \frac{3}{220} \theta^4 + \dots \right). \quad \dots\dots(43)$$

For small values of θ the magnification is thus only slightly less than in the space-charge-free case. The degree of magnification is quite small for angles less than $\pi/2$, less than three times, but for angles between $\pi/2$ and π it increases rapidly to many thousands of times.

§ 6. MOTION BETWEEN TWO COAXIAL RIGHT CIRCULAR CONICAL ELECTRODES WHOSE VERTICES COINCIDE

Consider first the space-charge-free case. Let (r, θ, ϕ) be spherical polar coordinates and let the axis of the cones be the line $\theta=0$. It is easy to show that the potential in the field between the cones can be written

$$V = a \ln \tan \frac{1}{2} \theta + b \quad \dots\dots(44)$$

where a and b are constants depending on the vertical angles and potentials of the cones. Clearly a possible form of the action function is $W = rg(\theta)$ and, as before,

$$g^2 + g'^2 = \frac{2e}{m} V. \quad \dots\dots(45)$$

The motion is confined to azimuthal planes and the equation to a trajectory has the form $r = r_0 F(\theta)$.

If the conical electrodes have the same vertical angle an extremely important property is to be noted, namely that the arrival pattern of the electrons on the anode

is a perfect magnification of the starting pattern at the cathode. This fact may be proved quite simply, for example by developing the two conical surfaces into a plane.

The integration of (45) with V as in (44) can be effected by a variety of means. For example, if α is the vertical angle of the first cone and if $t = \tan \frac{1}{2}(\theta - \alpha)$ and $T = \tan \frac{1}{2}\alpha$, a solution valid for $t < T$ can be obtained by expanding V as a power series in t and writing

$$g(\theta) = \sum_{n=1}^{\infty} a_n t^{2n+1/3}.$$

Proceeding in this way the equation to a trajectory becomes

$$r = r_0 \exp \frac{4}{3} t^2 \left[1 + \frac{1}{15} \left(\frac{1}{T} - T \right) t - \frac{1}{21} \left\{ \frac{32}{3} + \frac{11}{20} \left(\frac{1}{T} - T \right)^2 \right\} t^2 + \dots \right],$$

.....(46)

The property of perfect magnification also exists when the region between the cones is space-charge-saturated. This can be shown by the same argument as was applied in discussing motion between two inclined planes.

ACKNOWLEDGMENTS

The work was undertaken under a Turner and Newall Research Fellowship granted by the University of London.

The author is also indebted to Mr. A. M. Spooner for suggesting the problem of electron motion between inclined planes and for checking the numerical work.

REFERENCES

- MELTZER, B., 1949, *Proc. Phys. Soc. B*, **62**, 431, 813.
 STRATTON, J. A., 1940, *Electromagnetic Theory* (New York : McGraw Hill).
 SYNGE, J. L., 1937, *Geometrical Optics*, Cambridge Tracts No. 37 (Cambridge : University Press).
 WEATHERBURN, C. E., 1937, *Advanced Vector Analysis* (London : Bell).

A 100-Kilowatt Water-cooled Solenoid

By J. M. DANIELS

Clarendon Laboratory, Oxford

Communicated by F. E. Simon; MS. received 3rd July 1950

ABSTRACT. Design data and constructional details are given for a small water-cooled solenoid, which, for a consumption of 100 kw. produces a field of 14.7 kilogauss, uniform to within $\pm \frac{1}{2}\%$ in a cylindrical volume 6 cm. long and 4 cm. diameter. The solenoid contains no iron or other ferromagnetic material, and is in use at the Clarendon Laboratory, Oxford, for experiments on adiabatic demagnetization.

§ 1. INTRODUCTION

WE required for experiments on adiabatic demagnetization a solenoid capable of producing a strong magnetic field, if possible greater than 10 kilogauss. The nature of the experiments to be performed with it required that the field should be uniform to within $\pm \frac{1}{2}\%$ over a region some 6 cm. in length, that the solenoid should contain no iron or other ferromagnetic material, that the inside diameter should not be less than 57 mm. so that it would fit easily round a glass Dewar vessel of the type in use for these experiments, and that it should not be too bulky, and fairly easy to construct. We had at our disposal a d.c. generator which would deliver 300 amp. at 330 v. The solenoid to be described is one made to suit these requirements.

§ 2. GENERAL CONSIDERATIONS

The basic theory of high-powered air core solenoids was worked out long ago. Maxwell (1881) showed that, for a given available electric power, the greatest field obtainable from a uniformly wound coil is obtained when a longitudinal section through the coil, containing the axis of symmetry, exhibits the profile $r^2 = \sin \theta$. The line $\theta = 0$ is the axis of symmetry. This shape is useless for our purpose as the windings extend right to the origin, and it would be difficult to construct.

It has been shown that for coils with a rectangular section winding space (Fabry 1898, Cockcroft 1928) the field H at the centre is given by the formula

$$H = G \left(\frac{W\eta}{a_1\rho} \right)^{1/2}, \quad \dots\dots (1)$$

where W is the power dissipated in the coil, a_1 is the inner radius of the winding space, ρ is the resistivity of the winding material, G is a dimensionless factor which depends only on the shape of the winding space, and η is the 'filling factor' viz.

$$\frac{\text{volume of winding space occupied by conductor}}{\text{total volume of winding space}}$$

Fabry and Cockcroft both use a mixed system of units, expressing W in watts, a_1 in centimetres, ρ in ohm cm., and H in gauss. With these units the maximum value of G is 0.179 and for this the outer radius of the winding space is $3a_1$ and the axial length of the winding space $4a_1$.

Bitter (1936) considers the case of solenoids of other shapes than with rectangular section winding space and also with non-uniform windings. In this case.

equation (1) still holds. Values of G are given for these types of coil, many of which would be very difficult to construct. He gives, for the most efficient coil with any current distribution, $G = 0.272$. This coil extends to infinity in all directions. The best value of G quoted for any finite coil is a little greater than 0.2; coils with rectangular winding space and uniform winding do not compare unfavourably with these fancy-shaped coils.

A simple calculation shows that, for $a_1 = 3$ cm., about 50 kw. is required to produce 10 kg. in the most favourably shaped coil of rectangular section winding space. The heat so produced must be removed continuously; many methods have been suggested for cooling coils of this type, and are reviewed by Bitter (1936), Gaume (1947) and others. We decided to use a design originally due to Tsai (1947). Bare copper strip is used for the windings and adjacent turns are separated from each other by winding round the strip a nylon filament about 0.3 mm.

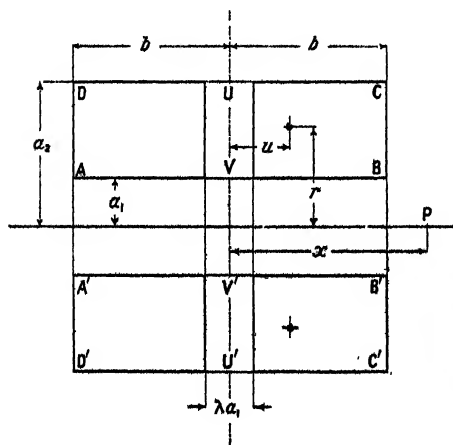


Figure 1. Longitudinal section through windings.

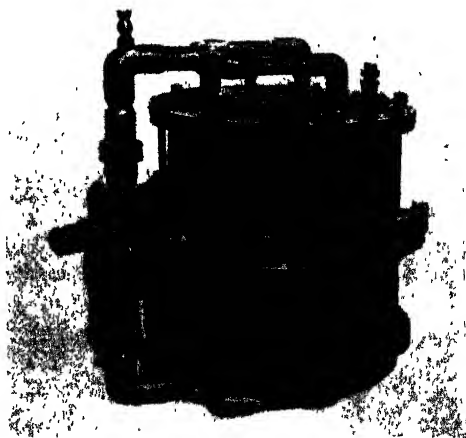


Figure 2. Photograph of the solenoid.

diameter, with a pitch of 3 mm. Ordinary tap water cannot be used as a cooling agent, as electrolysis takes place which rapidly corrodes the windings. Distilled water is therefore used, and cooled by tap water in a heat exchanger. This method of construction permits filling factors as high as 0.7, with high current densities and efficient cooling while requiring only small pressures to pass the water through sufficiently rapidly.

§ 3. DESIGN OF THE SOLENOID

In the first place it follows from Maxwell's electromagnetic equations that, if the field is uniform along the axis of the solenoid, it is also uniform away from this axis. It is, therefore, only necessary to ensure that the field is uniform along the axis. Now the very efficient coils mentioned in the previous section do not produce a very homogeneous field. For example, for the most efficient solenoid with rectangular winding space, the axial field has fallen by 1% at a distance $0.52a_1$ from the centre.

We decided to use a solenoid with rectangular section winding space, with length $6a_1$ and outer radius $3a_1$, and to omit from the centre of the coil a slice making a system analogous to the Helmholtz double coil (Figure 1). This was found to

give a sufficiently uniform field over the required volume: the value of G is 0.159, which is only 11% less than the theoretical maximum for a coil with rectangular section winding space and with no gap in the middle.

In order to calculate the shape of the winding space, it seemed to us an appropriate approximation (mainly because without it the calculations would have been far too tedious) to treat the windings as giving a uniform current density τ E.M.U./cm² flowing everywhere in the rectangle ABCD in Figure 1 and to account for the slice omitted by superimposing a surface current density $\lambda a_1 \tau$ E.M.U./cm. circulating in the plane annulus UV in a direction opposed to the direction of the main current. Thus λa_1 is the width of the gap.

If we consider the field at a point P on the axis of a circular current, radius r (Figure 1), the field is

$$\frac{2\pi r^2 i}{\{r^2 + (x-u)^2\}^{3/2}}.$$

Then for a solenoid with rectangular winding space

$$H_1 = \int_{u=-b}^{+b} \int_{r=a_1}^{a_2} \frac{2\pi r^2 \tau \, dr \, du}{\{r^2 + (x-u)^2\}^{3/2}}.$$

If we evaluate this and put $a_2 = \alpha a_1$, $b = \beta a_1$, $x = \xi a_1$, then

$$H_1 = 2\pi \tau a_1 \left[(\beta - \xi) \left\{ \sinh^{-1} \frac{\alpha}{|\beta - \xi|} - \sinh^{-1} \frac{1}{|\beta - \xi|} \right\} + (\beta + \xi) \left\{ \sinh^{-1} \frac{\alpha}{\beta + \xi} - \sinh^{-1} \frac{1}{\beta + \xi} \right\} \right]. \quad \dots (2)$$

Similarly, the flat coil lying in the plane UU' will produce a field

$$H_2 = \int_{a_1}^{\alpha a_1} \frac{2\pi r^2 (-\lambda a_1 \tau) \, dr}{\{r^2 + x^2\}^{3/2}}.$$

Therefore

$$H_2 = -2\pi \tau a_1 \lambda \left[\sinh^{-1} \frac{\alpha}{\xi} - \sinh^{-1} \frac{1}{\xi} - \frac{\alpha}{(\alpha^2 + \xi^2)^{1/2}} + \frac{1}{(1 + \xi^2)^{1/2}} \right]. \quad \dots (3)$$

If ξ is small, we get

$$H_2 \simeq -2\pi \tau a_1 \lambda \left[\ln \alpha - \frac{3}{4} \xi^2 \left(1 - \frac{1}{\alpha^2} \right) + O(\xi^4) \right]. \quad \dots (4)$$

The total field $H = H_1 + H_2$.

The axial field can be expressed as a Taylor series in terms of ξ and, by symmetry, only even powers of ξ will be present. In order to make the field uniform, we must arrange that as many as possible of the first terms in this expansion vanish.

Now

$$\left(\frac{d^2 H_1}{d\xi^2} \right)_{\xi=0} = 4\pi \tau a_1 \left[\frac{\beta^4}{(1 + \beta^2)^{3/2}} - \frac{\alpha^3 \beta^4}{(1 + \alpha^2 \beta^2)^{3/2}} \right]. \quad \dots (5)$$

This is always negative for a finite solenoid with no gap, so we must choose λ so that $(d^2 H/d\xi^2)_{\xi=0} = 0$.

Now, from (4), $(d^2 H_2/d\xi^2)_{\xi=0} = 3\pi \tau a_1 \lambda (1 - \alpha^{-2})$ so that

$$\left(\frac{d^2 H}{d\xi^2} \right)_{\xi=0} \propto 4 \left[\frac{\beta^4}{(1 + \beta^2)^{3/2}} - \frac{\alpha^3 \beta^4}{(1 + \alpha^2 \beta^2)^{3/2}} \right] + 3\lambda (1 - \alpha^{-2}) = 0.$$

This equation gives λ in terms of the known α and β .

If we desired, we could arrange two more slices equal in size and symmetrically placed about UU' to make $(d^4H/d\xi^4)_{\xi=0}$ zero also, and could continue this process of making further Taylor coefficients vanish.

In our case, where $\alpha = 3$ and $\beta = 3$, we find that $\lambda = 0.180$, justifying the assumption made at the beginning of this section that the slice omitted is thin. For the coil constructed, we put $\lambda = 0.25$. This gives two small humps in the graph of H against ξ , one each side of the origin, and further extends the region in which H is constant to within $\pm \frac{1}{2}\%$ (see Figure 3).

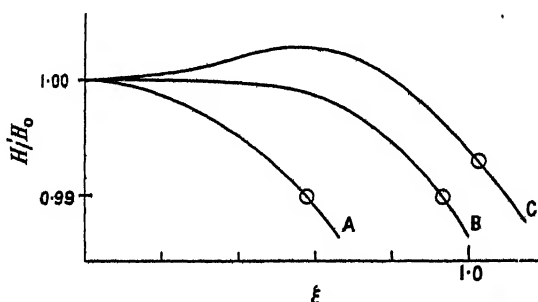


Figure 3. H/H_0 as a function of ξ . A: for a coil $\alpha=3$, $\beta=3$, no gap; B: $\alpha=3$, $\beta=3$, $\lambda=0.182$; C: $\alpha=3$, $\beta=3$, $\lambda=0.250$. The circles denote the end of the useful region, i.e. where the total variation in H is less than 1% of H_0 .

The final step in the design is to choose the copper strip to give a total resistance equal to the optimum load resistance of the generator which will supply the power; in our case this was 1.1 ohms.

§ 4. CONSTRUCTION

The coil consists of 20 flat coils, or 'pancakes', wound with copper strip $\frac{1}{4}$ in. \times 20 s.w.g. The inner end of the strip was soldered on to a thin brass tube 61.5 mm. outside diameter, $\frac{1}{4}$ in. long, and $\frac{1}{2}$ mm. thick. The solder used was Johnson Matthey's L. M. 10. Nylon thread 0.3 mm. diameter was wound round the copper strip by means of the winding device shown in Figure 4.

The tube A which carries a cheek on the left is a push fit into the ball race B and into the sprocket S, and is turned by an electric motor, which drives S by means of a chain. The ball race B is clamped into a thick aluminium bracket C. D is a bobbin which rotates freely on A, and on to which the nylon filament is first wound. The copper strip XX passes along the axis of A, and moves from right to left during the winding. It is located centrally by passing through a slot in the plate E, which is held loosely in a circular groove in the end cheek of A, so that E stays still as A rotates. The end cheek of A also carries an arm, on the end of which is a wheel F. Nylon filament from the bobbin D passes over the wheel F and thence round the copper strip. The copper strip was fed into the device at a speed of about 2 ft/min, from a pair of rollers, driven from the same motor as that which drives the sprocket S. By a suitable choice of reduction gear between the motor and the rollers, the pitch of the nylon winding could be adjusted to the desired value. This winder was designed and made by one of our mechanics, Mr. J. J. Milligan, who also attended to many of the finer details, not mentioned here, which ensure its smooth working.

The distance between consecutive turns of nylon was about $2\frac{1}{2}$ mm. The brass tube was slipped on a bush between two cheeks $\frac{1}{4}$ in. + 0.6 mm. apart, and the pancake was wound at the same time as the nylon was wound on the strip. When the required outer diameter for the pancake, 181.5 mm., was reached, the copper strip was severed and the pancake temporarily secured by wrapping insulating tape round its circumference. The pancakes were insulated from each other by means of wheels cut from tufnol or paxolin sheet, with 8 spokes, 2 mm. wide, shaped so as to impede the water flow as little as possible. These wheels were cut from $1\frac{1}{2}$ mm. sheet, except the specially thick wheels used in the centre to provide the gap in the winding, and at the ends to provide collecting spaces for the cooling water.

When designing electrical conditions between the pancakes, it must be remembered that the contact resistance must be kept very small. In addition, equation (1) shows that it is important to have the inside radius of the windings as small as possible, and the following construction was therefore used. For the inside connections, thin brass tubes were made, each $\frac{1}{2}$ mm. thick, $\frac{1}{2}$ in. long, with a stiff

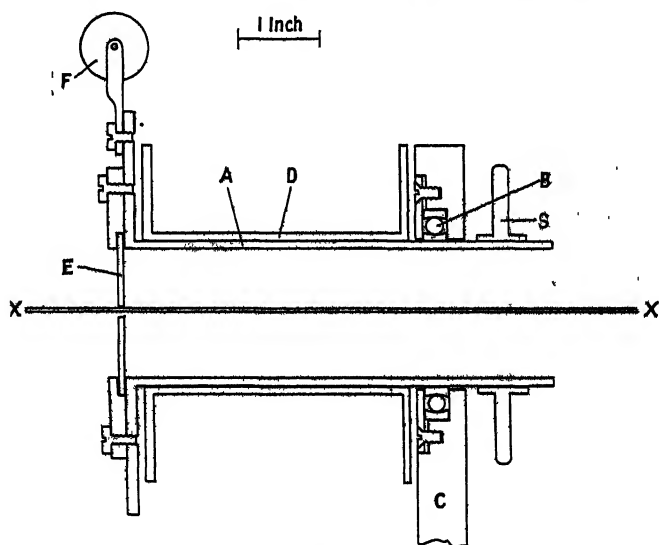


Figure 4.

push fit into the tubes on which the pancakes were wound. The pancakes were placed together in pairs, with a paxolin wheel between each, the diameter of these wheels being just less than the inner diameter of the outer case. The electrical and mechanical connection between the two pancakes was made by pressing one of these brass tubes into the hole in the centre. The inside of the magnet case was a brass cylinder, 59 mm. outside diameter, 16 s.w.g. thick and 22 cm. long. A few layers of polythene* sheet were wrapped round this, and the pairs of pancakes were slipped on with a paxolin wheel between each pair, the outer diameter of these wheels being just less than the outer diameter of the pancakes. Electrical connection was made between adjacent pairs of pancakes on the outside, by removing the temporary binding of insulating tape, and winding a large copper band round both adjacent pancakes. The ends of the band were secured by two screw clips, one

* Originally, a sheet of resin-bonded cloth was used, but this broke down and a spark passed between the windings and the case. This occurred after a few weeks' use, and is believed to have been due to water entering the material via the ends of the cloth fibres.

bearing down on each pancake. Special bands were used for the centre and for the ends, and the terminals were soldered to the bands securing the end pancakes.

The outer cylindrical case was a copper cylinder 8 in. outside diameter, and 10 s.w.g. thick. The large paxolin wheels have an outside diameter which allows them to slide into this, leaving just sufficient room for a sheet of paxolin paper between their edges and the outer case. The end cheeks are made of tufnol $\frac{5}{8}$ in. thick. These are held together by eight long brass rods, $\frac{1}{4}$ in. diameter tapped 0 BA at the ends. The joins between the end cheeks and the rest of the case, and where terminals and water-pipes pass through the end cheeks, are sealed by means of 'Gaco' rubber rings, in seals whose construction is obvious from the diagram (Figure 5). The cooling water enters at three points spaced 120° apart

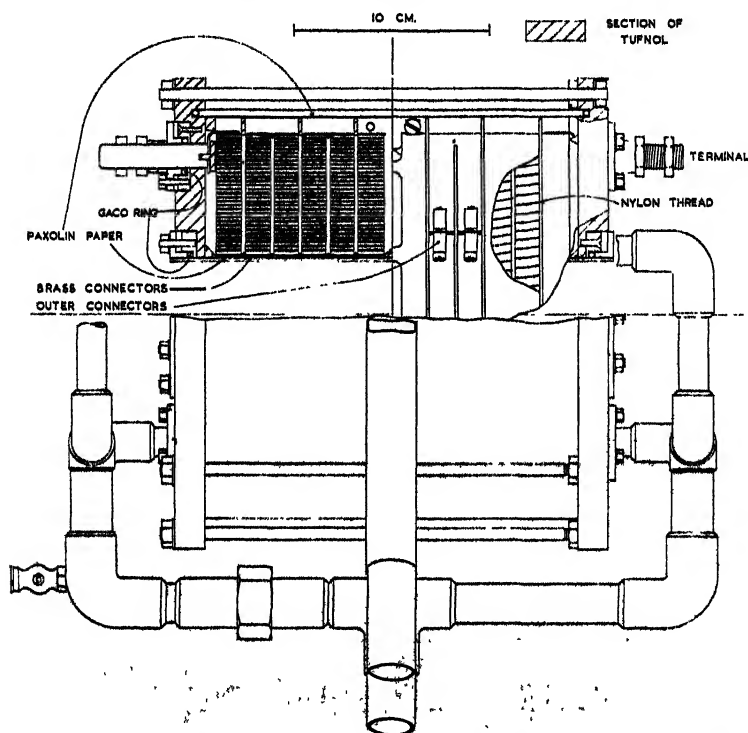


Figure 5. Section and elevations showing construction of the solenoid. For simplicity in this diagram the solenoid is shown as being composed of 12 pancakes.

at the middle of the copper outer case and leaves at three equally spaced points on each end cheek. Holes are cut in the paxolin paper lining of the case, and in the central copper connecting band, to let the water through. An air release valve of the type used on radiators for central-heating systems is provided to ensure that no air is trapped in the solenoid.

§ 5. PERFORMANCE

The cooling water is circulated by a 'Safran' centrifugal water-pump, made by the Saunders Valve Company, which delivers about 40 gal/min. against a pressure of 3 atmospheres. This water is cooled with tap water in a heat exchanger, similar to the one described by Hudson (1949). The field at the centre of the solenoid is linear with current, and with 280 amps a field of 14.7 kilogauss is obtained. The

filling factor for the solenoid is 0.43, and it is easily verified that the field given by equation (1) agrees with the experimental value; with this current the power consumed is 99.5 kw. The resistance of the coil when cold is 1.16 ohms and the temperature rise of the cooling water on passing through the solenoid is about 8° c. The weight of the solenoid and case, without cooling water, is 67 lb.

The homogeneity of the field was measured, using two identical search coils connected in opposition, fixed on a rod so that their separation could be varied. When removed from the field, the throw of the ballistic galvanometer to which they were connected was proportional to the difference in the fields at the two coils. Within the limits of experimental error, the results are in agreement with the calculated values of the field.

This solenoid has given many hours of reliable service in this laboratory, with the exception of one breakdown mentioned in a previous footnote. Work is in progress on the design and construction of another coil for our 2,000 kw. generator.

ACKNOWLEDGMENTS

The author expresses his thanks to the Department of Scientific and Industrial Research for a maintenance grant, and also to Dr. N. Kurti for helpful advice. He is also indebted to Dr. Swallow of Imperial Chemical Industries (Plastics Division), for the free gift of the nylon monofilament used for insulating the windings.

REFERENCES

- BITTER, F., 1936, *Rev. Sci. Instrum.*, **7**, 482.
 COCKCROFT, J. D., 1928, *Phil. Trans. A*, **227**, 317.
 FABRY, C., 1898, *Eclairage Electrique*, **17**, 133.
 GAUME, F., 1947, *J. Recherches du Centre National de la Recherche Scientifique*, **3**.
 HUDSON, R. P., 1949, *J. Sci. Instrum.*, **26**, 401.
 MAXWELL, J. C., 1881, *Treatise on Electricity and Magnetism*, 2nd Ed. Vol. II (Oxford: Clarendon Press), p. 331.
 TSAI, B., 1947, *Report of an International Conference on Fundamental Particles and Low Temperatures*, Vol. II (London: Physical Society), p. 89.

LETTERS TO THE EDITOR

Precision Lattice-Parameter Measurements

Recent measurements by the present author (Keith 1950) of the lattice-parameters of one natural and three synthetic samples of clear crystalline quartz have given results which differ from one another by more than the experimental error and which, except for one of the synthetic samples, are significantly smaller than the accurate values given by Lipson and Wilson (1941). The absolute precision of the measurements, as distinct from the precision relative to a specific camera calibration, was slightly uncertain, but believed to be of the order of 1 part in 10⁵. In view of the criticism based upon these results of the quartz calibration standard for x-ray powder cameras it seemed desirable to confirm this accuracy estimate, and hence the disagreement between the results and the values of Lipson and Wilson. This has now been done by comparing the parameters previously measured for a sample of Brazilian quartz with a 19 cm. 'Unicam' camera with those determined for the same sample with a 6.4 cm. Straumanis camera (Straumanis 1949) which has recently been constructed in this laboratory. Cameras of the Straumanis type are absolute instruments in which an asymmetrical film arrangement is employed to circumvent the necessity for preliminary calibrations. Copper K α radiation was again used, and the calculation of the

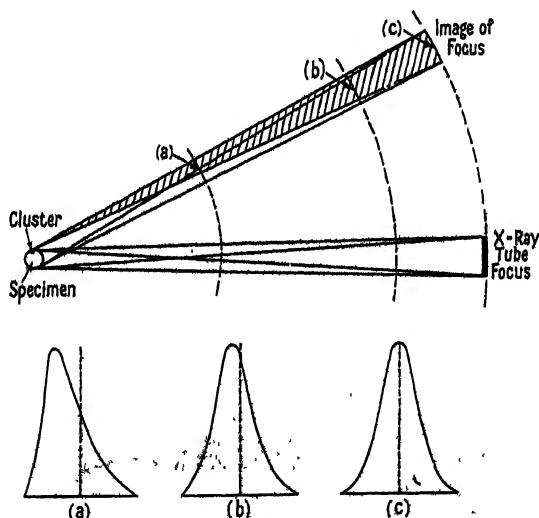
quartz parameters from the measured Bragg angles for four separate films was carried out exactly as in the previous experiments. The results are compared below with those obtained earlier (values in kx).

19 cm. 'Unicam' camera (25°c.) $a=4.90309_3$; $c=5.39367_3$; $c/a=1.100055$.

6.4 cm. Straumanis camera (25°c.) $a=4.90314_4$; $c=5.39373_3$; $c/a=1.100056$.

It is seen that the two sets of values agree to within 1 part in 9×10^4 , showing that the absolute precision of the earlier parameters is almost as great as was estimated. Comparison of values of the lattice-parameter of silver bromide measured with these two cameras has further confirmed the accuracy of the determinations carried out with the 'Unicam' camera. It is interesting to note that Straumanis has obtained a consistent absolute precision of 1 part in 2×10^5 in measurements of the lattice-parameter of Al which he has carried out with his cameras.

Some observations which have been made during the investigation of quartz with the Straumanis camera are worthy of mention. With a specimen 0.25 mm. in diameter slightly spotty diffraction lines were obtained at all angles, and for the experiments a 0.5 mm. diameter specimen had to be used to ensure uniform diffraction lines. Even taking into account the small height (1 mm.) of the specimen being irradiated this behaviour is surprising, and is explained principally by the unfavourable ratio of the specimen-film and specimen-target distances which was used. The effect of varying this ratio is clarified by the accompanying diagram, showing in exaggerated form the influence on diffraction-line contours of a



Diffraction-line contours at various film positions (see text).

cluster of crystallites reflecting strongly at any given instant from one localized region of the specimen. Film position (a) corresponds to the situation in the Straumanis camera, and position (b) to that in the 'Unicam' camera. A considerable improvement would be brought about by shortening the collimator of the Straumanis camera and advancing the specimen towards the target. Powder photographs of metal systems taken with this camera are generally free from spottiness, even with thin specimens, due to the high multiplicity factors of the reflecting planes for simple structures and the inherent diffuseness of the x-ray reflections from metallic crystals.

In conclusion, it should be emphasized that the above effect may, in the case of crystals giving sharp x-ray reflections, invalidate the use of very thin specimens which are often desirable (Keith 1950). The unsuspected lack of uniformity which may occur in the line contours of apparently perfect diffraction lines would be most easily overlooked with the high measuring microscope magnifications also recommended previously. Provided that due care is taken and, in particular, that the ratio of the specimen-film and specimen-target

distances is kept as near as possible to unity, the advantages of thin specimens and high magnifications are, nevertheless, considerable. It is conceivable that some of the parameters reported earlier for synthetic quartz may contain slight errors due to this effect and the high magnification, which was used. The validity of the general conclusion regarding the reliability of camera calibrations by the quartz method cannot, however, be affected.

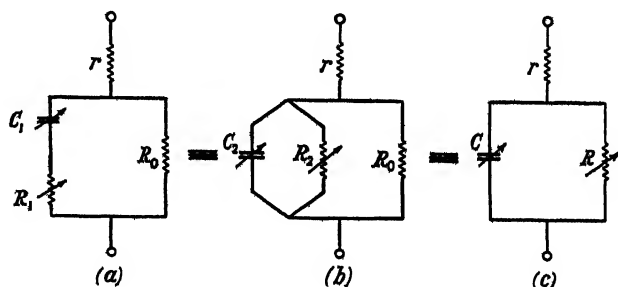
H. H. WILLS Physical Laboratory,
University of Bristol.
4th September 1950.

H. D. KEITH.

KEITH, H. D., 1950, *Proc. Phys. Soc. B*, **63**, 208.
LIPSON, H., and WILSON, A. J. C., 1941, *Proc. Phys. Soc.*, **53**, 245.
STRAUMANIS, M. E., 1949, *J. Appl. Phys.*, **20**, 726.

The Capacitance of Selenium Rectifiers

In his recent paper R. Cooper (1950) suggested that the observed variation of impedance of a selenium rectifier in the audio-frequency range showed the presence of anomalous dispersion. However, from his Figure 3 (in which the ordinate axis should be labelled as reactance) this conclusion does not appear to be unequivocal since a good fit to *all* the points can be obtained by a circle whose centre does not lie on the resistance axis but at approximately (1300, -200). Such a circular locus is applicable to an impedance which can be represented in either of the forms shown in (a) and (b) in the diagram which are equivalent to the form (c), the symbol of variability implying frequency dependence. The



circular locus for impedance is obtained when $\omega C_1 R_1$ or $\omega C_2 R_2$ is a constant independent of frequency. In the consolidated form (c) it is seen that both the components C , R representing the barrier layer impedance are in general functions of frequency. The property of a circular locus is readily seen by noting that $C_1 R_1$ or $C_2 R_2$ in parallel with R_0 has a straight-line locus in the admittance plane which becomes a circle by inversion into the impedance plane; the resistance r then displaces the circle by r to the right in the impedance plane.

Furthermore, by applying the relationship between the real and imaginary parts of an impedance to $C_1 R_1$ or $C_2 R_2$ it is found that on writing

$$\omega C_1 R_1 = \frac{1}{\omega C_2 R_2} = \tan \frac{m\pi}{2} \quad 0 \leq m < 1$$

the capacitances C_1 or C_2 must vary as ω^{-m} and the resistances R_1 or R_2 as ω^{-1+m} .

This behaviour is of the type discussed by Fricke (1932) in connection with electrolytic polarization assuming a relaxation function of the form t^{-m} . It has been observed experimentally in selenium barrier-layer photocells by Wood (1933) who made impedance measurements on the audio-frequency range. It is not surprising that there should be an analogy between the time lag behind the applied field of the polarization and depolarization at an electrode in an electrolyte and of the ionization and de-ionization of impurity atoms in the barrier layer of a rectifier. In the case of selenium with a cadmium counter-electrode

the activation energy is probably of the order of 0.1–0.2 eV., which is larger than $k\theta_D$ (where θ_D is the Debye temperature of selenium), and the theory of Goodman, Lawson and Schiff (1947) then indicates relaxation frequencies of the order of those observed by Cooper.

Radio Research Station,
Slough, Bucks.
15th August 1950.

R. E. BURGESS.

COOPER, R., 1950, *Proc. Phys. Soc. B*, 63, 176.

FRICKEL, H., 1932, *Phil. Mag.*, 14, 310.

GOODMAN, B., LAWSON, A. W., and SCHIFF, L. I., 1947, *Phys. Rev.*, 71, 191.

WOOD, L. A., 1933, *Rev. Sci. Instrum.*, 4, 434.

The Correction of the Spherical Aberration of Electron Lenses using a Correcting Foil Element

Several systems for the correction of the spherical aberration of electron lenses have been proposed (Scherzer 1947, Gabor 1949, Hubert 1949, Ramberg 1949) but have all been shown to provoke practical difficulties which have precluded their use, with the possible exception of the crossed cylindrical lens system proposed by Scherzer.

In particular, it has been established that a departure from field continuity, by the introduction of conducting foils into the optically active region of an electron lens, may lead to a correction of the spherical aberration (Scherzer 1949). The performance of a suitable foil correction element, which is attractive because of its basic simplicity, has been examined in this laboratory. The characteristics of this element are outlined in the following.

Consider first a plane parallel foil element, with foils separated by a distance d and maintained at potentials ϕ_0 and ϕ_1 , which produces a uniform retarding field in an intermediate region between the object Q_0 and the objective lens (Figure 1). In this field the

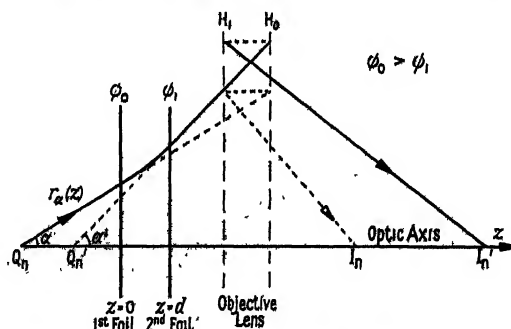


Figure 1. The uniform retarding field produced by a plane parallel foil element deviates an electron trajectory $r_\alpha(z)$ away from the axis, in such a way as to correct the spherical aberration of a succeeding objective lens, which is represented by its principal planes H_1 and H_2 .

radial component of velocity of an extra-axial electron trajectory $r_\alpha(z)$ is unaffected, whereas the axial component of velocity is diminished, and a divergence of the electron trajectory results. In addition, it is apparent that this divergence is less for a paraxial trajectory, because of the smaller radial component of velocity. Thus the foil element produces a negative spherical aberration which will correct the positive aberration of the following objective lens. This simple physical interpretation is confirmed by a rigorous analysis.

The trajectory of an electron, which originates with inclination α from an axial object point, which is taken in coincidence with the first foil, may be expressed

$$r(z) = -\frac{\sin 2\alpha}{2K} \left[1 - \left(1 + \frac{2Kz}{\cos^2 \alpha} \right)^{1/2} \right],$$

$$K = (\phi_1 - \phi_0)/2\phi_0 d.$$

It follows that the effective displacement of a 'marginal object point' is

$$\Delta Q_n = d - \frac{\cos^2 \alpha}{K} \left(1 + \frac{2Kd}{\cos^2 \alpha} \right)^{1/2} \left[\left(1 + \frac{2Kd}{\cos^2 \alpha} \right)^{1/2} - 1 \right],$$

so that the resulting displacement between 'marginal' and 'paraxial' object points may be expressed, in third-order approximation,

$$\Delta_n = -\frac{d}{4} \frac{\Delta \phi}{\phi_0} \tan^2 \alpha \left[1 - \frac{1}{2} \frac{\Delta \phi}{\phi_0} \frac{(1 + \cos^2 \alpha)}{\cos^2 \alpha} \right]$$

$$(\Delta \phi / \phi_0)^2 < 1, \quad \Delta \phi = \phi_1 - \phi_0.$$

Accordingly the condition for the correction of the third-order spherical aberration of the main lens becomes

$$\frac{\Delta \phi}{\phi_0} = -\frac{1}{2} \left[\left(1 + \frac{16C_s}{d} \right)^{1/2} - 1 \right],$$

where C_s is the spherical aberration constant (Glaser 1940 a).

An estimate of the practical requirements for correction is obtained by using values of the aberration characteristics C_s/f for the axial magnetic lens distribution,

$$H(0, z) = H(0, 0) / [1 + (z/a)^2],$$

tabulated by Glaser (1940 b). It is found that the correction of a typical lens requires a foil separation d equal to the focal length f of the main lens, with a potential increment $\Delta \phi = -0.7\phi_0$. Thus the conditions for correction, as given by this simple treatment, appear not unreasonable.

Similar reasoning shows that a foil element which introduces an accelerating field after the main lens (Figure 2) will also correct the spherical aberration. However, in this case,

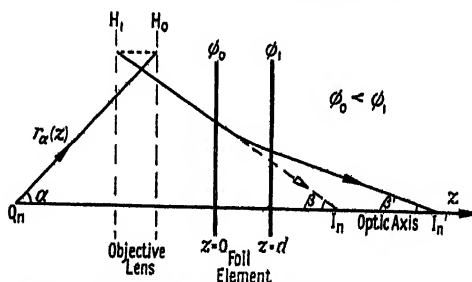


Figure 2. The accelerating foil element deviates an electron trajectory $r_\alpha(z)$ away from the axis in such a way as to correct the spherical aberration of a preceding electron lens.

the deviation of the imaging electrons that the element must produce is increased in proportion to the magnification of the objective lens, and the appropriate foil potentials are prohibitive.

The simple analysis given above serves to indicate the many problems arising in the practical design of a foil-corrected lens. The more important of these problems will now be mentioned.

In the corrected lens a part of the foil element will need to be located within the main lens, if a reasonable stage magnification is to be obtained. Thus a discrete treatment of the performance of the foil element is not fully comprehensive. In addition, the superposition of the lens and element fields will result in an electron retardation within the lens, with an accompanying modification of the electron-optical performance of the main lens, and it will be necessary to introduce overall changes in the main lens construction to allow for this effect.

Very thin beryllium foils can be obtained (Hast 1948) which may prove suitable for use in correcting the spherical aberration. However, it is probable that spurious effects, due to scattering of the electrons within the foil and to the charging of surface contamination, may invalidate the corrective properties of the foils. This difficulty may be surmounted arbitrarily. The object may be mounted on the first foil itself. Further, it is suggested that an apertured electrode may replace the second foil without appreciably disturbing the

field configuration, provided that the aperture is small compared with the foil separation. However, a complete treatment of this problem will need to account for the finite field perturbation produced by the aperture, and must include the analysis of the superimposed action of the lens and element fields.

Further investigation of these problems is being continued in this laboratory.

The writer has recently been informed, through private communication, that Mr. M. E. Haine, of the Associated Electrical Industries Research Laboratories, has carried out a similar investigation in an analysis based on the analogous negative spherical aberration of a parallel-sided glass plate (Dyson 1949).

The writer wishes to express thanks to Professor J. Sayers for his encouragement throughout this work, and to Dr. G. Liebmann for his constructive criticism of this corrective system. This work is made possible by a grant from the Department of Scientific and Industrial Research.

Department of Electron Physics,
University of Birmingham.

U. F. GIANOLA.

MS. originally submitted 16th June 1950;
shortened to Letter 8th September 1950.

DYSON, J., 1949 *Proc. Phys. Soc. B*, **62**, 565.

GABOR, D., 1949, *The Electron Microscope*, 2nd Edition (London: 'Electronic Engineering').

GLASER, W., 1940 a, *Z. Phys.*, **116**, 19; 1940 b, *Ibid.*, **117**, 285.

HAST, N., 1948, *Nature, Lond.*, **162**, 892.

HUBERT, P., 1949, *C. R. Acad. Sci., Paris*, **228**, 233.

RAMBERG, E. G., 1949, *J. Appl. Phys.*, **20**, 183.

SCHERZER, O., 1947, *Optik*, **2**, 114; 1949, *J. Appl. Phys.*, **20**, 20.

Aurora and Luminous Night Clouds

At about 2200 G.M.T. on 24th July 1950 there appeared together above the northern horizon in Central Scotland an auroral arc with summit base at an elevation of about 10° and luminous night (noctilucent) clouds reaching an elevation at 5° (Plate I). The simultaneous occurrence of these two phenomena must be very rare indeed. As twilight waned, the clouds extended steadily over the north-eastern sky and by 0050 G.M.T. on the 25th, they covered an area reaching along the horizon between azimuths 345° and 85° and upwards to a maximum elevation of nearly 25° . Meanwhile the aurora had vanished or had been extinguished by the brilliant light of the clouds. The wave structure, characteristic of these clouds, is shown in Plate II.

Using a Störmer-Krogness auroral camera, photographs showing cloud development were taken at regular intervals during the night until the clouds disappeared against the intensifying background of skylight at 0309 G.M.T. (sunrise 0402 G.M.T.). Unfortunately, photographs were obtained from only one station out of four normally used for simultaneous photography, but, as previous observation has shown that the clouds are always situated at about 85 km. (Jesse 1896, Störmer 1933 a, b, 1934, 1935 a, b), even single photographs allow their position, rate of movement and wave-crest separation to be measured if this altitude is assumed.

The constitution of luminous night clouds is uncertain. That their luminosity comes from sunlight scattered by the cloud particles is established, for their spectrum shows the Fraunhofer lines but no emission lines, and the predominant colour is blue. (During this display, a change in colour of portions of the cloud from vivid blue to white was observed, suggesting growth of the particles. However, no such change had been observed during a fine though less extensive display on the night of 10-11 July 1949 (Paton 1949).) The appearance of the clouds always at a height of between 80 and 90 km. accords with the height distribution of temperature calculated by Whipple (1943) from meteor track photographs. These reveal a large lapse rate of temperature between a maximum of 350° K. at 60 km. and a minimum of about 200° K. below an inversion at 80 km.

Vegard (1935) has suggested that the clouds are formed by the condensation of water vapour on sodium oxide nuclei, formed by the combination of sodium and atmospheric oxygen. But it is difficult to explain how a supply of water vapour sufficient in quantity to

form extensive clouds may exist at this great height, though Vegard proposes that it might be formed by the combination of atmospheric oxygen, perhaps in the form of ozone, and hydrogen, precipitated into the atmosphere from the sun. However at 80 km., where a temperature minimum exists and the pressure is of the same order of magnitude as the vapour pressure of water, conditions might conceivably be favourable for condensation.

The more general belief is that the clouds consist of residual meteoric or comet dust, which under the prevailing temperature distribution drifts to and remains below the base of the inversion at about 80 km. This is the view held by Professor Störmer who has witnessed and made photographic measurements of many displays.

Because the clouds appeared very frequently during the summers following the Krakatoa eruption in 1883 (the clouds are visible only during the summer months and between latitudes 45° and 62°), it was at first thought that they consisted of the products of condensation of gases projected to great heights during a volcanic eruption. But Vestine (1934) has pointed out that during the period 1880–1887 there occurred a remarkable number of meteor appearances, showers and great comets, and that this evidence, together with the fact that brilliant displays followed immediately after the great Siberian meteor of 1907 while no conspicuous occurrence succeeded the eruption of Katmai, Alaska, in 1912, strongly favours a cosmic origin of the cloud particles.

An analysis of the results of the measurements of the present plates will be published elsewhere.

Department of Natural Philosophy,
University of Edinburgh.
2nd October 1950.

J. PATON.

JESSE, O., 1896, *Astr. Nachr.*, **140**, 161.

PATON, J., 1949, *Met. Mag.*, **78**, 354.

STÖRMER, C., 1933 a, *Publ. Univ. Obs. Oslo*, No. 6; 1933 b, *Avk. norske Vidensk. Akad., Oslo*, No. 2; 1934, *Nature, Lond.*, **134**, 219; 1935 a, *Astrophys. Norv.*, **1**, 87; 1935 b, *Nature, Lond.*, **135**, 103.

VEGARD, L., 1935, *Geofys. Publ.*, **10**, 53.

VESTINE, E. H., 1934, *J. Roy. Astr. Soc., Can.*, **28**, 249, 303.

WHIPPLE, F. L., 1943, *Rev. Mod. Phys.*, **15**, 246.

High Frequency Permeability of Ferromagnetic Materials

With reference to the paper by Millership and Webster (1950), it seems desirable to mention some later work which bears on the generality and precision of the results there given.

All the results given show a monotonic decrease of the inductive permeability, μ_L , with increase of frequency, although more recent experiments on nickel carried out in this department indicate that there are definite deviations from this simple monotonic form, showing that this must not be regarded as general. It is hoped that these results will be published when completed.

In connection with the precision it should be stated that each point on the graphs in the above-mentioned paper represents the mean of a group of experimental values. Re-determination of the inductive permeability for annealed iron, using improved apparatus which gives a much reduced scatter of the individual experimental points, has shown a somewhat more rapid decrease in the 150 to 250 Mc/s. range than is shown in Figure 5 of the paper. The general conclusions, however, remain unaffected by the more precise results obtained in this new determination.

Physics Department,
The University, Leeds, 2.
1st November 1950.

F. V. WEBSTER.
K. S. DRIVER.

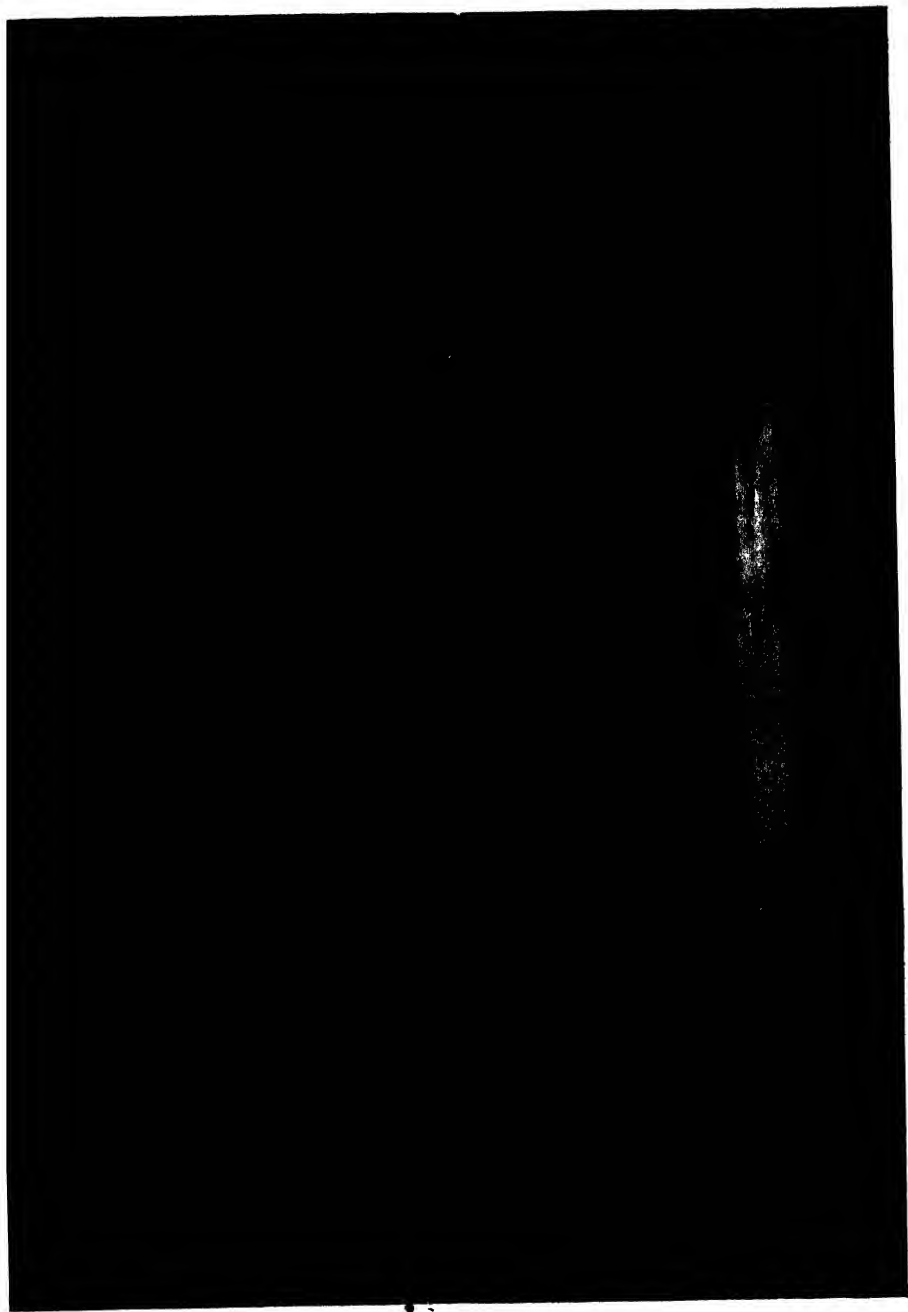


PLATE I.

Auroral arc with sunlit ray and luminous night clouds just above the horizon.
Photographed at Abernethy, Perthshire, 2355 G.M.T., 24th July 1950.



PLATE II.

The luminous night clouds at 0148 G.M.T. on 25th July 1950, showing characteristic wave structure.

Obituary Notices

JAMES REGINALD ASHWORTH

THE death of James Reginald Ashworth on 9th July 1950 deprived scientific circles in the north-west and farther afield of a highly esteemed member. At the age of eighty-nine, he had only recently resigned from the office of honorary secretary of the Rochdale Literary and Scientific Society after sixty-five years, and he had been a co-opted member of the Public Libraries, Art Gallery and Museum Committee of Rochdale for sixty-three years.

Dr. Ashworth built up a considerable reputation in the wider scientific world. He wrote eighteen books and pamphlets dealing mainly with various aspects of magnetism. In May 1939 he was invited to address a conference on magnetism organized at Strasbourg by the French Central Council of Scientific Research.

For many years he had been deeply interested in the subject of atmospheric pollution and its effect on the transmission of ultra-violet radiation. He installed a gauge on the roof of the Rochdale Technical School and made daily observations to the Smoke Abatement Society.

At the time of his death Dr. Ashworth was curator of the Manchester Literary and Philosophical Society. Up to the end of last year he was in good health, attending the Council meetings regularly and giving valuable advice.

He had been a member of the Physical Society for fifty-three years and was also a Fellow of the Royal Meteorological Society.

The following tribute was paid to Dr. Ashworth by the vicar of Rochdale (the Rev. G. E. N. Molesworth) :—

“He said he valued above all things integrity and truthfulness of heart and mind. So indeed his life has proved. We shall not see his like again and we shall be immeasurably the poorer for it. He has reached the end honourably with his finger on the pulse of passing events.”

ALFRED JOSEPH BULL

DR. ALFRED JOSEPH BULL was an outstanding personality of our time. Few men have gained distinction and honour in such widely different spheres of thought and action as geology and the graphic arts.

He started his career as an analytical chemist and later joined the staff of Messrs. Lumière where he was one of the pioneers in colour photography. He continued research on colour vision at the Northampton Institute where he lectured on Optics. In 1903 he joined the staff of the School of Photo-engraving and Lithography at the old house in Bolt Court, once a residence of Dr. Johnson. Here with A. J. Newton, the Principal, he investigated the problems of colour photography in relation to photo-engraving, especially with regard to the sensitivity of photographic emulsions, the transmission of filters, and the testing of lenses for colour work.

In 1912 Bull became Principal of the new school built on the original site in Bolt Court. He approached his new responsibilities from a new angle. Research was not regarded as a proper activity for a technical school by the Education Authority of that time. Despite this, Bull concentrated on technical research to such a degree that it became the educational tradition of Bolt Court and helped to establish its international reputation. His own contribution from the year 1912 to the time of his retirement in 1946 was considerable. His special field was colour reproduction, and his work formed a foundation for much of the development of ‘colour masking’ in recent years. Alone or in collaboration with his staff he published numerous papers on other aspects of photo-engraving including tone-rendering, densitometry, and ‘pattern’ in colour reproduction.

In 1933 Bull was elected President of the Royal Photographic Society and continued in office to play a leading part in the Fox Talbot centenary the following year. During the last war he was concerned with research for the War Office, and Field Survey, and on techniques for the production of ‘graticules’. The story of this work and of the establishment of munitions factories at Bolt Court and Putney has been told in a report he published at the end of the war and in a series of papers read at the Royal Photographic Society.

Bull's geological work covered a wide field including geomorphology, stratigraphy and tectonics. He became a member of the Geologists' Association in 1916, served on its Council for many periods, became President, 1926-38, and subsequently served as Vice-President and successively as Secretary of the Fields Meetings and Publications Committees. As Chairman of the Weald Research Committee from its inception he was active in carrying out an important part of the geological survey of the South Downs, the river valleys and the South Coast.

He was also a Fellow of the Geological Society and for some time Treasurer of the Mineralogical Society. As a result of several visits to the Alps with Swiss geologists he became interested in tectonics and contributed valuable papers on the theory of Mountain Building. In this work he carried out many interesting laboratory experiments, and his papers were usually illustrated with photographs he had taken in the field.

His personal friends will remember Bull as a cheerful and interesting companion, always happiest when pursuing his geological interests. In retirement he continued much of his committee work and he was following his geological studies with vigour up to the time of his death. He was a modest man seeking only to find the truth in all that he did. Nothing could have pleased him more than the knowledge that the work he had started in such diverse fields would be continued in the same spirit of sincerity.

H. M. C.

FREDERICK CHARLES CLARKE

THE death of Mr. F. C. Clarke, a Fellow of the Physical Society since 1911, occurred after a very short illness on 16th February at the age of 72. His death will be deeply regretted by many old friends.

He was a native of Plymouth, having a brilliant scholastic career as Foundation Scholar of the Plymouth Grammar School. From there he proceeded with a National Scholarship to the Royal College of Science in 1898 and gained the Associateship of the College in Physics, and the Honours Degree in Physics of the London University.

Clarke was deeply interested in the growing activities of the many large and well-equipped Technical Colleges, and he acquired unique teaching experience by holding appointments at the Birkbeck College, Chelsea Polytechnic, and as teacher of Mathematics and Physics at the West Ham Technical College, founding courses which were recognized for Internal Students of the London University. In this way Clarke gained exceptional experience in studying the requirements of students both for University work and for work in Technical and Applied Physics, and thereby qualified himself for his subsequent career.

The outbreak of the war 1914 to 1918 removed him temporarily for military duties, as he was an active member of the Corps of London Electrical Engineers, R.E.(T). He was actually in camp at Dover commanding a Searchlight Company at the outbreak of hostilities, with the rank of Captain R.E.(T), and subsequently assumed command of the Dover searchlights with the rank of Major throughout the war.

After the war he resumed his duties as Lecturer in Physics at West Ham until 1930, when he was appointed as the first Principal of the Rotherham College of Technology, a position he held until 1948. His capacity was so highly appreciated that his term of service was extended five years beyond the normal retiring age. During that time Clarke devoted all his energies to equipping laboratories, recruiting staff, contacting the employers of the great steel and other industries of the neighbourhood, and the social life of the students of the College.

On his retirement many tributes were paid to the zeal with which he fulfilled his duties. The amazing growth of the College was stated as being undoubtedly due in no small way to his tremendous drive and to the intimate relations he held with his staff and students. The original staff of four increased to 42 full-time and 88 part-time teachers, and the number of students increased from 750 to over 3,000. The courses included all branches of engineering, from the principles of management and administration to the duties of foremen and supervisors in local works. Students worked for University Degrees in electrical engineering and metallurgy, chemistry, physics and mathematics.

Clarke kept in close touch with all developments in Technical Education, both local and national. He was past-president of the Association of Teachers in Technical Institutions, and later a member of Council of the Association of Principals of Technical

Institutions. He was also a member of Council of the Sheffield Section of the Institute of Production Engineers, and of the Sheffield Subsection of the Institution of Electrical Engineers, of which he was an Associate Member.

Probably visits to the Physical Society Exhibitions gave him the most pleasure, for here he could meet old friends and watch the wonderful achievements of those, who unlike himself, could devote themselves to physics research: he was at heart a physicist.

After a bare eighteen months of retirement he undoubtedly paid the penalty of the strain of a very active life.

He leaves a widow, a son and a daughter.

W. S. T.

JAMES ARNOLD CROWTHER

THE sudden death of Professor J. A. Crowther on 25th March at the comparatively early age of 66 came as a great shock to his many friends, and is a severe loss to the Physical Society, of which he had been a distinguished member for nearly 23 years. Since his retirement in 1946 he had lived in Cornwall, and although for some time his health had not been good his zest for life remained, and to the last he was actively interested in literary, scientific and educational matters. When I last saw him in December 1949 there was no sign of any flagging of his spirits, and in a letter which he wrote to me a few days before he died he sounded characteristically cheerful and happy. Our deepest sympathy is due to his widow and his two sons in their time of sorrow.

James Arnold Crowther was born in Sheffield on 28th August 1883, and received his early education at the Royal Grammar School. From there he entered St. John's College, Cambridge, with an open Science Scholarship, and after gaining First Class Honours in both parts of the Natural Sciences Tripos he started work as a research student in the Cavendish Laboratory under J. J. Thomson. He held the Mackinnon Studentship of the Royal Society, and in due course became a Fellow of his College. Most of the investigations going on in the Cavendish at that time were connected with ionizing radiations, and so it came about that Crowther entered the field of x-rays and radioactivity, which was to become the chief scientific interest of his life, and to which he made a number of important contributions. Later he became a pioneer in the study of the effects of penetrating radiations on living cells, and may fairly be said to have laid the physical foundations of radio-biology.

Crowther early showed exceptional gifts as a teacher. His style was lucid and always interesting, whilst his understanding of human nature, combined with a keen sense of humour, enabled him to keep his audience alert and eager to hear what was coming next. After serving an apprenticeship as a demonstrator under Dr. G. F. C. Searle, he was put in charge of the electrical laboratory, and also gave stimulating lectures on Electricity and Magnetism to students working for Part I of the Natural Sciences Tripos. When the Cambridge Diploma in Radiology was started in 1919, Crowther was appointed to the new University Lectureship in Physics as applied to Medical Radiology, a position which he held for four years, and in which he did splendid service in teaching comparatively advanced physics to medical graduates whose previous knowledge of the subject was scanty, and whose interests were mainly clinical. He was particularly well qualified to do this as, during the 1914-18 war, he had worked in the x-ray department of a war-time hospital, and had helped to develop radiographic technique at a time when all the work was done by means of gas tubes, and only the most rudimentary forms of equipment were available.

In 1924 Crowther was appointed to the Chair of Physics in the University of Reading, where he threw himself energetically into the task of organizing the physics teaching and building up a school of research. He was the author of a number of successful books, of which the best known is his *Ions, Electrons and Ionising Radiations*. This book has passed through many editions since it first appeared in 1919, and has probably been read by every honours student of physics throughout the Empire during the past 30 years. It was no small task to keep a work of this character up to date, for the issue of a new edition had to be followed almost immediately by preparation for the next. His *Manual of Physics*, also first published in 1919, still retains its popularity, whilst the *Handbook of Industrial Radiology*, of which he was editor, has passed through two editions in five years. In his earliest work, *Molecular Physics*, published in 1914, as much as in his later writings,

he showed the same clarity of style and the same gifts for arousing interest that characterized his lectures. Little more than a year ago he gave a most successful broadcast on his old Professor, J. J. Thomson, for whom he had always a profound admiration and affection. This talk brought him letters of appreciation from many old Cavendish men who shared his deep regard for 'J. J.' Crowther had a very wide experience as an examiner, and was as much interested in the proper conduct of School and Higher Certificate examinations as in special honours B.Sc. work.

Crowther served on the Council of the British Institute of Radiology for a number of years, was President in 1936-37, and gave the Silvanus Thompson Memorial Lecture in 1937. In 1947 the Institute elected him to Honorary Membership, and he was also an Honorary Member of the Faculty of Radiologists. He was a Founder Member of the Institute of Physics, of which he was Honorary Secretary from 1932-46, Vice-President in 1948, and to which he gave devoted service. He did much to raise the status of physics as a profession, and during the second World War he served for two years as Vice-Chairman of the Parliamentary and Scientific Committee, which was concerned with the proper use of scientific manpower.

Crowther's married life was a very happy one, and he owed much to the loyalty and devotion of his wife. Those who were privileged to enter his home could not fail to be struck by the warmth and friendliness of its atmosphere. His conversation was always amusing, and he had a ready wit, but he was essentially a man of simple tastes, who enjoyed a game of tennis or golf, and was always ready in the evening to make up a four at bridge. He took pleasure in foreign travel and spent many happy holidays abroad with his wife and family touring Europe in his car. More recently he had formed the wish to visit New Zealand to see his son who is engaged in academic work there. He had some gifts as a singer, and when he attended the annual Cavendish Dinner he could always be relied upon to give a tuneful and spirited rendering of the special Cavendish songs, most of which were written by the late A. A. Robb. He was particularly fond of the Gilbert and Sullivan operas, which he knew almost by heart. He had complete gramophone records of many of them, and never tired of playing them over in the evenings, particularly if he had a sympathetic and appreciative audience. His early association with the Cavendish Laboratory, as well as his later work at the Institute of Physics, brought him into intimate contact with large numbers of physicists, who will always remember him as much for his human qualities and personal charm as for his notable scientific achievements. He had a great gift for friendship, and will be sadly missed by all who knew him. G. STEAD.

GEOFFREY E. F. FERTEL *

THE tragic death of Geoffrey Fertel on 19th January 1949 marked the end of a man of remarkable ingenuity and individuality, great integrity, and yet considerable diffidence.

He was born on 19th November 1913, and his father was organist at Bromley Parish Church. He was educated at Sevenoaks School, and later came to study physics at the Royal College of Science. Here his almost prodigious experimental ability developed to the full and, not content with the experiments provided in the degree course, he devised and carried through a great deal of practical work at his home, including even an absolute determination of the ohm and the construction of Geiger counters and the necessary auxiliary equipment. All this was done with materials and components costing at most a few shillings, obtained from the most diverse places. He was a rigid functionalist and, especially at this time, took a kind of inverted pride in the fact that the non-essential parts of his apparatus were never unnecessarily good. Although he devoted most of his energy to practical physics and took little part in games or any other social activities, his interests were by no means as narrow as one might have thought on first meeting him. One discovered after a time that he was enviably well read in literature, both prose and poetry, and something of an authority on wild plants. It was at this time too that his practical joking was at its maximum—he had several ways of producing loud explosions at unexpected times and places, and one lecturer may perhaps remember spending an embarrassing time trying to bring the slide lantern into operation, only to find at length that one of the carbons was insulated by a tightly fitting black paper tube.

* This notice was unfortunately just too late for publication in the 1949 volume of the *Proceedings*.

After graduating he worked under Sir George Thomson on neutrons, much of his time being spent with a team working on a time-of-flight experiment. Shortly before the war he left London to take up a post at the University of Bristol, where he did some work with Professor C. F. Powell on the photographic plate technique of detecting nuclear particles, then still in an early stage of development. On the outbreak of war he joined a mobile group fitting up coastal radar stations, and shortly after went on a cruise to the Mediterranean on one of H.M. ships to carry out radar trials. On such a retiring and introspective person these turbulent experiences had a profound effect. For the rest of the war he worked in Admiralty laboratories on radar, first at Portsmouth, later at Nutbourne, and finally back at the Physics department of Bristol University, where an Admiralty team under R. W. Sutton was working on the development of new valves. He specialized in gas discharge devices for common aerial working, and it is probably in this field that his most important work was done. At the end of the war the valve group moved from Bristol, but he remained there, now back on the University staff, and spent some time teaching and applying his knowledge of radar technique to more academic problems.

He spent most week-ends exploring the countryside on foot or bicycle, and it was at Bristol that he developed an interest in caving and canoeing. He delighted in any sort of climbing, and had an alarming technique of proceeding upside down, feet first, which was particularly effective when applied to trees. He will be remembered affectionately by everyone in the Physics Department at Bristol for the Heath Robinson devices which he delighted in making for Christmas parties. One of the best of these was a cigarette-lighter of arresting proportions. Built on the stand of an old astronomical telescope, and having a large and crooked tree-trunk as a prominent part of its structure, it attained a height of seven or eight feet, and was set in operation by first winding up some heavy pieces of scrap iron to a considerable height. On pulling a lever these weights were released and drove a primitive dynamo, the output from which heated a small wire. In the meantime the cigarette, which had been inserted in an ornate holder made from an old curly brass candlestick, was brought into contact with the wire, and had applied to its other end a partial vacuum produced by means of a bicycle pump, also driven by the weights. The probability of successful operation of the lighter was about the same as that of the more conventional kind. Apart from diversions of this kind he was an excellent though unconventional mechanic, and throughout his career made most of his apparatus himself, usually with remarkable speed.

In 1948 he obtained leave of absence from Bristol University to join the team working on the cyclotron at Birmingham University, under Professor Oliphant. By this time he had lost much of his former reserve and made many new friends.

The precise nature of the accident which resulted in his electrocution can never be known, since he was by himself in the cyclotron pit at the time, but there is no doubt that he died instantly.

The amount of published work he has left hardly does justice to his creative ability. Technical discussion with him was never easy: his mind was always leaping ahead, or perhaps sideways, but always to fresh ground, and whether or not any definite conclusion had been reached at the end, there were always dozens of new ideas and original suggestions for overcoming practical difficulties. All who knew him will agree that his personality was unique, and the contribution he had to make to science, not only by his own work but by the energy and enthusiasm which he could inspire in others, can ill be spared.

D. F. GIBBS.

CYRIL OWEN GREEN

CYRIL OWEN GREEN was born on 24th November 1920 and was educated at Campbell Square School, Northampton. He next studied as an evening student at Northampton College of Technology and after obtaining the Higher National Certificate in Electrical Engineering decided to study Physics. He entered University College, Nottingham, in October 1942 with a Foundation Scholarship. After a short course Green's war service was spent as an electrical officer in the Fleet Air Arm.

Green returned to Nottingham in 1945 and obtained first class Honours in Physics in the London B.Sc. examination in July 1948. He then became a research student at

Manchester University working on the properties of oxide-coated cathodes. Later, in 1949, he joined the cosmic-ray group directed by Professor P. M. S. Blackett. He completed a useful analysis of the problem of ionization measurements in the Wilson cloud chamber.

Last autumn Green spent three months helping to erect a large electromagnet at the observatory on the Pic-du-Midi in the Pyrenees. In May 1950 he again went to the Pyrenees in order to install a Wilson cloud-chamber in the field of the magnet. On 11th May, a colleague, and three French workers at the observatory stated to climb the Pic. They were travelling light, made good progress and were resting at 8,000 ft. when Green suddenly collapsed and died shortly afterwards. The other members of the party and the observatory staff rendered what help they could, but it was of no avail.

In this way a career of great promise was brought to an early end. Green made many friends who will remember his quiet unassuming manner and unselfish disposition. In 1948 Green married Miss Lorna Rumney.

N. DAVY.

HERMAN SHAW

THE sudden death of Dr. Herman Shaw, Director of the Science Museum, at his home at Barnes on 4th May 1950, came as a surprise and shock to his many friends in the Museums and scientific worlds and in particular to the Physical Society, of which he was Honorary Treasurer. His record as administrator, scientist and man of affairs was an impressive one, and his friendliness and accessibility had endeared him to a very wide circle of acquaintances.

He was born on 14th October 1891, the only son of the late Mr. G. H. Shaw of Huddersfield. He was educated at Bradford Grammar School, where he obtained a Governors' scholarship, and at the Imperial College of Science and Technology, which he entered as a Royal Scholar in 1911. In 1913 he won an Aeronautics Research Scholarship at this College. In the first World War he became a Lieutenant in the R.N.V.R. in 1915, a Flight Lieutenant in the R.N.A.S. in 1916, and a Captain in the R.A.F. in 1917.

Shaw entered the Science Museum in 1920, and for the remaining 30 years of his life his energies were devoted to the development and expansion of its collections and services, and to research in Applied Geophysics, in which subject he became a world authority.

His work in Geophysics began in 1921, when he and the late E. Lancaster-Jones obtained permission to make extensive laboratory tests on an Eötvös torsion balance which had just been acquired by the Museum. The tests were described in two papers published in the Physical Society's *Proceedings*, and proved so successful that they were soon extended to field experiments. These were carried out first at a site in Shropshire and then in 1925 at Hodbarrow, Cumberland, using a British-made torsion balance in the design of which Shaw and Lancaster-Jones played a leading part. The experiments, the first of their kind to be carried out in the British Isles, showed in a striking way that an underground rock structure—in this case a limestone 'dome'—could be located and delineated with the aid of the torsion balance.

In 1928 Shaw and Lancaster-Jones designed their 'gravity gradiometer', a form of torsion balance specially modified for field work in the measurement of gravity gradients, and much smaller, lighter and quicker in operation than the earlier Eötvös form of balance.

Shaw's interest in Applied Geophysics was by no means confined to the gravity method of prospecting, but soon grew to cover all its aspects, and he became recognized as one of the leading authorities on the subject, on which he published in all about 20 papers.

Concurrently with his geophysical work, most of which was carried out in his spare time, Shaw was very active in purely Museum work. The ten years following World War I was a busy period for the Museum, for they saw the completion and occupation of the new East Block, which was formally opened by H.M. King George V in 1928. Considerable expansion took place in the Science Collections, and in this development Shaw played his full part. He was also responsible to a large extent for the special exhibition on Applied Geophysics held at the Museum in 1931, and was co-author of the Museum Handbook on the subject published at this time.

He was made a Deputy Keeper in 1931 and in 1935 was promoted Keeper of the Department of Physics and Geophysics. In 1939 he was put in charge of A.R.P. arrangements at the Museum, and later took full charge of the arrangements for the evacuation

to the country of the greater portion of the Collections. This work he carried out with great efficiency and thoroughness, and it was largely due to him that the Collections came through the Second World War with a minimum of damage or deterioration.

Soon after the outbreak of the War, the Director of the Museum, Col. E. E. B. Mackintosh, was called away for military duties, and in 1940 Shaw was made Acting Director. Col. Mackintosh later returned to the Museum, but on his retirement in November 1945, Shaw was appointed to succeed him as Director.

He resolved at once that the Museum should be opened again to the public at the earliest possible date, and applied himself energetically to the task. Partial re-opening took place as early as February 1946, and Shaw's policy received immediate justification by the fact that by the end of that year over $1\frac{1}{2}$ million visitors had already seen the limited collections then on view. Over the succeeding four years steady progress was made in rehabilitation, and at the time of his death the work was nearing completion, although somewhat restricted in total extent by the loss of the Old Building through age and war damage.

In spite of the burden imposed by this work, Shaw found time to play a notable part in many outside activities. His association with the Physical Society in particular was a long and intimate one. He was elected a Fellow in 1914, and was a member of Council from 1935 to 1939 and from 1941 to 1946, in which year he was elected Honorary Treasurer, a position which he held at the time of his death. He was also a member of the Colour and the Optical Groups. He was a Founder Fellow of the Institute of Physics.

At the time of his death Shaw was also President of the Museums Association, a trustee of the Imperial War Museum, a governor of the Imperial College of Science and Technology and a manager of the Royal Institution.

Shaw married, in 1919, Constance, the daughter of Mr. F. Shaw of Harrow. They had one son, who died as a young man in 1946, after an illness involving much suffering near its close. The loss was a severe blow to Shaw and his wife, but he bore it with fortitude, and only those in close contact with him were able occasionally to realize what it had cost him.

Amid such an active life, Shaw left himself little time for hobbies. He was, however, proud of his garden, and took considerable pleasure in motoring.

As a man, his most notable characteristics were his energy and approachability. He never spared himself in work, but obviously enjoyed the wide range of contacts which it brought him. As Director he was just, firm and kindly, and showed a practical interest in the well-being of all his staff. His sudden death at the height of his powers is a severe blow to the Museum to which he devoted so much of his energy and talents, and will be widely regretted among his large circle of acquaintances in many walks of life.

F. A. B. W.

GEORGE WILLIAM TODD

GEORGE W. TODD, Professor of Experimental Physics in King's College, Newcastle-upon-Tyne, died on 24th February 1950, a few weeks after an operation. His health had been causing his friends anxiety for many months, but he had carried on with his work, although with growing difficulty, until the end of the previous term.

He was born in Birmingham in 1886 and, after attending a secondary school in that city, entered the University in 1904, where he studied under Poynting and graduated with Honours in 1907. After two years' research at Birmingham he was elected to an 1851 Exhibition, which he held at the Cavendish Laboratory under Sir J. J. Thomson. This was extended beyond the normal period of two years until 1912, when he took up a science mastership at the Wattville Road School in Birmingham. In the following year he was awarded the D.Sc., and was appointed to the Royal Grammar School, Newcastle-upon-Tyne, as Physics Master. Here he remained until 1916, when he joined the Munitions Inventions Department of the Ministry of Munitions, later becoming head of the Physics Section. At the end of the war he was appointed to the newly established Chair of Experimental Physics at Armstrong College, Newcastle, where he remained until his death.

His research work was mainly experimental and covered a wide field. It was characterized throughout by elegance, accuracy and thoroughness, for he possessed in full measure the

true physicist's appreciation of the aesthetic aspect of his subject. These qualities were first shown in his 1909 paper on the thermal conductivity of air and other gases, and they were equally apparent in his 1928 paper on an expansion method of measuring the Peltier coefficient. His sojourn at the Cavendish Laboratory led to the publication of four papers on ionic mobilities, and provided him with a fund of pleasant reminiscence from which both he and his friends derived much enjoyment. It was evidently the most stimulating and satisfying period of his scientific career.

His scientific work during the 1914-18 war was on quite different lines, being mainly concerned with gas reaction velocities, and on this subject he published in all some six papers. The most important of these was probably that in which, in collaboration with S. P. Owen, he derived expressions for the temperature coefficients of the velocity and equilibrium constants of homogeneous gas reactions. A little later he showed that physical equilibrium could be treated in a similar manner, and so obtained a vapour pressure equation which gave good agreement with experiment over a wide range of temperature. In the same year (1920) he published the results of some work on the rate of flow of gases through capillary tubes, from which he attempted, with a considerable measure of success, to establish a formula which should be applicable down to the lowest pressures.

At the time of his appointment to Armstrong College, and for some years later, research facilities in the Physics Department were very meagre and the teaching duties heavy. He published several further papers, and prepared a new and revised edition of Poynting and Thomson's *Properties of Matter*, but devoted himself mainly to lecturing, both inside and outside the College. In this he was most successful, having the gift of clear and simple exposition and a sympathetic appreciation of the difficulties of even the least intelligent of his audience. These attributes, together with his kindly and genial attitude to his students, earned him their esteem and affection, and his death was keenly felt by them. His colleagues also held him in the highest regard, for his sincerity, good-nature and unselfishness were patent to all. During the second war, when half the staff, including myself, were away, he took charge of the department and shouldered a very heavy burden of teaching and administrative work for nearly six years. This was entirely typical of him. He never sought responsibility, but was always ready to take on extra work for the benefit of others.

His main interests outside physics were gardening and painting. He was for many years a member of the Newcastle Society of Artists, and for the last few years of his life its Chairman. He was a Founder Fellow of the Institute of Physics, a member of the Faraday Society, a Fellow of the Physical Society since 1922, and Editor of the Durham University Philosophical Society's *Proceedings* since 1921. He gave numerous extra-mural courses and lectures, which were invariably well attended, and served as a member of the Investigating Panel for Higher Certificate Examinations in 1924-6, and again in 1937-8.

He will be remembered not only as a physicist who made many and diverse contributions to knowledge and as an unusually gifted expositor of his subject, but also as a man of most kindly and lovable nature. He will be sadly missed by his innumerable friends.

W. E. CURTIS.

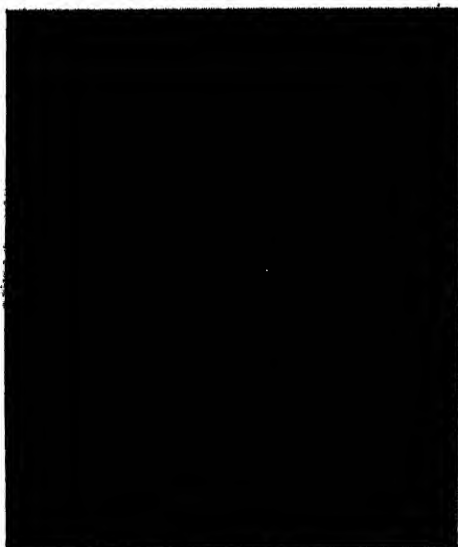
ALEXANDER WOOD

The late Dr. Alexander Wood, M.A., D.Sc., Fellow and Tutor of Emmanuel College and Lecturer in Experimental Physics in the University of Cambridge: "I have no specialist's contribution to make. My rôle has always been that of interpreter rather than the research worker, and I only take courage from the fact that never was the rôle of interpreter more essential than in our own time." Thus Wood spoke of himself in his inaugural address to the Acoustics Group of the Physical Society (March 1947). He modestly left it there. But there is something more to be said.

Why with his intelligence and force was he, in these times, an academic scientist only. He was born in 1879 and died in 1950. He came to Cambridge from Lord Kelvin's degree lectures at Glasgow University just at the right time. He had the passion for enquiry. He was vitally interested in the scientific arena and was historian of the Cavendish Laboratory; yet in research had no contribution to make. I believe he deliberately left the laboratory, that he left it because, though enquiry was great, there was in fact something greater, and that his significance lies in the fact that in the considerable stress of our times he would not



HERMAN SHAW



JOSEPH ALFRED BULL



JAMES ARNOLD CROWTHER

give up either orthodox Christianity on the one hand nor modern science on the other. Of course it is not surprising that he was also religious. In all periods there have been acute scientists who have seen nothing contrary to good sense in associating a Creation with a Creator and who would turn from sense data *in simplici intuitu*. But with Wood there was (in my opinion) a distinguishable factor. He was of the blood and the temper of the Scottish Reformers. He had the instinct of the theologian to organize the whole of knowledge, and felt that one half—Creation—is ultimately unintelligible without the other half. For him Truth was the aim and the battlefield, but Truth meant 'the Whole'. And it is natural that in his own historical period the battlefield must include the physical sciences because in that direction, in response to a profound gaze of man's faculties (what the mystics call a *quality of attention*), new knowledge and new power was emerging. He well knew the excitement of the chase—when detective plus inductive powers are in full cry—and always insisted that the creative faculties were as fully involved. But where was it leading, and into what theological pattern could it fit? Once early in the thirties we discussed this in the Fellows' Garden at Emmanuel. The debate in Christian terms went somewhat thus: "Autonomous enquiry is inevitable—to know is a passion. But God permits scientists, endows them with infinite curiosity; then he permits ultimate risks, certainly as part of the contingency of free-will and a cosmic design not fool-proof. What then becomes of the Kingdom of God on earth? Here is the dialectic of Christianity: on the one hand the City of God, the reformer's inspiration, on the other the eschatology of the New Testament, the facing of catastrophe; religion is the technique of being at home on the edge of the precipice. It is for this reason that Christianity is so distasteful to Reason but so true as an account of Life itself." But Wood did not agree. When the deeps opened under him in the course of argument he had a peculiar metaphysical smile known to his friends—a smile of exhilaration. No, Christian Truth must include cosmic Truth. He meditated much on the Gospels and saw Our Lord before Pilate—"For this cause came I into the world, that I should bear witness to the truth". Wood writes: "The induction upon which we are called upon to act is not merely that the good life is the life lived and taught by Jesus: it is the whole view of God and the world from which his life issued and in virtue of which the life was and is possible." In his later years as nuclear physics developed, the conviction of an all-in Christian Truth was not easy. Deeply anxious, he would look out over his peaceful Broads, or scan the rows of new Cambridge houses which, as a mathematician and architect, he had done so much to build, and the words 'radioactive mud' would frame themselves. Yet he did not despair. There was always something to be done in the immediate battlefield. He could preach, argue, defend his religion, "but first he followed it himself". His action came out of his inner life. So he allowed himself to be many-sided, yielding to the service of God now here, now there. Like many pioneers he was a boisterous fighter, and in committee was a master of the tactical battle. The range of his activities was astonishing. In addition to his duties as fellow and tutor of his college, and then as chairman of the Borough Housing Committee, there were many others. He was a member of the Cambridge Union (the first speaker at Swanwick), the Peace Pledge Union, the Central Nations Good Organisation, the Cambridge Divisional Labour Party, the Welfare of Youth Union, the Fellowship of Reconciliation, and many others besides.

But action was not merely an anodyne. It was the testing of theory: it was part of his experimental method. Science was always much more than enquiry. It was also the selecting of what to enquire into, and it was the organizing and unifying of our knowledge. "We are face to face with an amazing multiplicity and diversity of phenomena constituting our physical environment. If we are to live intelligently we must organize our past experience so as to make it readily available for meeting the circumstances of the moment and for providing for the needs of the future. In the earlier stages the results of past experience are accumulated instinctively as an unorganized background in our minds. Later we become conscious of the need for organization, and Science is the response to this very practical need." He then continues: "... in a sense religion is also a response to a similar practical need. Man's experience is not confined to sense data. He has to live in a world which, on the face of it at least, is a moral and even a spiritual world as well. Just as successful living in the first of these worlds—or the first plane in the only world we know—involves a mastery of our material environment, so successful living in the world's wider aspects means mastery in the moral and spiritual realms also. There is no more reason for us to be the playthings of our moral environment than there is for us to be the sport of our

physical environment." The test of both, then, is successful living. "We want to understand life. We want to discover the great induction which will relate and give meaning to all its complex and puzzling experiences." He points to the fact that in Science it is not, at any particular time, the fashionable 'verification' which truly convinces us of a theory; it is the fact that the theory over a long period has contributed to render phenomena intelligible. And if it does so all kinds of anomalies and contradictions are endured. "Newton's Law of Gravitation has been subjected to innumerable verifications. Yet it seems probable that gravitation is a property of space, and not, as Newton thought, of matter." The importance of 'verification' is, however, that it shall make intelligible our future experience also. "This also we may fairly demand of religion. Our lives ought to be a continuous verification. We must not only understand the situations which have arisen in our past experience; we must be equipped to meet new situations as they arise. This double criterion is undoubtedly our test of scientific truth, and I conceive it to be the *intellectual* test of religious truth as well." (*Italics are his.*)

Wood thus gives philosophical significance to 'successful action' in the total issue of theoretical knowledge, and he transfers it from Science to Religion. Hence the importance of his active life. He was, moreover, fully conscious of the metaphysical weakness of parallels drawn from sense data. He believed there was no such thing as a logical refutation of philosophical idealism. Our conviction of the reality of things arises from constantly living among them and acting as if they were real. But if we direct our attention to, if we gaze at, material things only, their sense of reality grows because our powers of concentration are almost demonic in their summoning power. But by the same test spiritual things ignored lose their sense of reality. "In a world where the things of sense and time press on us more continuously and more aggressively than ever before, the reality of the unseen can only remain an unshakable conviction if we allow it to make its own impact upon us, and that means a scrupulous anxiety to preserve times and seasons when alone, or in fellowship, we may have leisure for worship and meditation and prayer. All these times must be opportunities for the unseen to make its own impact, and not times devoted to a fussy activity of our own. They must include spaces of silence in which God can speak to us. As we enter the doorway of religion it is simplicity that matters most. If we happen to be scientists or theologians we must not be disloyal to the spirit and temper and results of our scholarship, but neither must we wait outside in the vain hope that all our difficulties will be resolved."

The bulk of the quotations given above are taken from his book *In Pursuit of Truth—A Comparative Study in Science and Religion* (S.C.M. Press, 1927) and from *Some Implications of Modern Physics* (Independent Press Ltd., 1936).

H. BAGENAL.

CONTENTS FOR SECTION A

	PAGE
Dr. J. G. DAUNT and Dr. K. MENDELSSOHN. Film Transfer in Helium II: I—The Thermo-Mechanical Effect	1305
Dr. J. B. BROWN and Dr. K. MENDELSSOHN. Film Transfer in Helium II: II—Influence of Geometrical Form and Temperature Gradient	1312
Mr. R. BOWERS and Dr. K. MENDELSSOHN. Film Transfer in Helium II: III—Influence of Radiation and Impurities	1318
Dr. K. MENDELSSOHN and Dr. G. K. WHITE. Film Transfer in Helium II: IV—The Transfer Rate on Glass and Metals	1328
Dr. B. H. BRANSDEN and Dr. E. H. S. BURHOP. The Disintegration of the Deuteron by Neutron Impact	1337
Mr. E. W. FULLER. A Coincidence Absorption Study of the Decay of ^{251}Hf	1348
Mr. H. MESSEL and Dr. D. M. RITSON. On the Development of the Nucleon Component of the Cosmic Radiation in Air	1359
Letters to the Editor:	
Dr. R. J. BENZIE and Dr. A. H. COOKE. The Magnetic Susceptibility of Copper Caesium Sulphate	1366
Mr. A. C. HOLLIS-HALLETT. Oscillating Disc Experiments in Liquid Helium II	1367
Dr. B. BLEANEY and Mr. H. E. D. SCOVIL. Nuclear Spins of Neodymium 143 and 145	1369
Mr. J. K. MACKENZIE. The Stresses and Energies associated with Inter-Crystalline Boundaries	1370
Corrigendum	1371
Obituary Notices:	
JAMES REGINALD ASHWORTH	1372
ALFRED JOSEPH BULL	1372
FREDERICK CHARLES CLARKE	1373
JAMES ARNOLD CROWTHER	1374
GEOFFREY E. F. FERTEL	1375
CYRIL OWEN GREEN	1376
HERMAN SHAW	1377
GEORGE WILLIAM TODD	1378
ALEXANDER WOOD	1379
Contents for Section B	1382
Abstracts for Section B	1382
Subject Index, Section A, Vol. 63	1385
Index of Authors (with Titles), Section A, Vol. 63	1392
Index to Reviews of Books, Section A, Vol. 63	1399

ABSTRACTS FOR SECTION A

Film Transfer in Helium II: I—The Thermo-Mechanical effect, by J. G. DAUNT and K. MENDELSSOHN.

ABSTRACT. The existence of a thermo-mechanical pressure difference has been established with the transfer film instead of a capillary connecting the two volumes of liquid helium II. It is shown that under these circumstances reversible conditions are closely approximated since return flow of heat or liquid through the film is negligible. The speed with which a thermo-mechanical pressure difference can be established is limited by the same critical transfer rate as that determining film transport under gravitation.

Film Transfer in Helium II: II—Influence of Geometrical Form and Temperature Gradient, by J. B. BROWN and K. MENDELSSOHN.

ABSTRACT. Experiments have been carried out in order to investigate a possible dependence of the film transfer in liquid helium II on the geometrical conditions of the measuring arrangement or on a temperature gradient along the film. Using transfer vessels of different shape, no variation in the rate of transfer could be detected. Equally no change from the ordinary rate of flow could be found under temperature differences up to 3.5×10^{-2} deg.

Film Transfer in Helium II: III—Influence of Radiation and Impurities, by
R. LOWERS and K. MENDELSSOHN.

ABSTRACT. In contradiction to the results of de Haas and van den Berg, the complete exclusion of thermal radiation has not been found to increase the transfer rate beyond the normal value. It was observed, however, that contamination of the transfer surface with condensed gases will cause a greatly increased film flow. The effect of impurities on the transfer has been studied in detail and the conclusion has been reached that the high and variable transfer values recently reported by some observers are to be ascribed to this cause. The bearing of the effect on cryogenic technique is discussed.

Film Transfer in Helium II: IV—The Transfer Rate on Glass and Metals, by
K. MENDELSSOHN and G. K. WHITE.

ABSTRACT. The rate of transfer R has been measured as a function of temperature along surfaces of glass, platinum and nickel. While the (R, T) curves have all approximately the same shape, the absolute values of R on metals were found to be much higher than on glass. The effects of polishing and degassing of the transfer surfaces have been studied. Possible explanations for the observed effects are discussed.

The Disintegration of the Deuteron by Neutron Impact, by B. H. BRANSDEN and
E. H. S. BURHOP.

ABSTRACT. The method of distorted waves is applied consistently to the calculation of the cross section for disintegration of deuterons by neutrons. At an incident neutron energy of 11.5 Mev. the cross section calculated for the process (6.4×10^{-28} cm²) is about one-third that obtained in previous calculations but is nevertheless an order of magnitude greater than that indicated by the rather meagre experimental data. The calculated cross section is also larger than the maximum permitted by considerations of the conservation laws. This indicates that the n - d interaction is so strong that the requirements for the validity of the distorted wave method are not satisfied.

At an incident neutron energy of 5.1 Mev. where it is considerably less than the permitted maximum, the calculated cross section is of the same order of magnitude as that deduced indirectly from measurements of the p - d disintegration cross section, allowing for the effects of the electrostatic repulsion between the proton and the deuteron.

A Coincidence Absorption Study of the Decay of ^{181}Hf , by E. W. FULLER.

ABSTRACT. The decay of 46 day ^{181}Hf has been investigated by absorption coincidence techniques using an integral type delayed coincidence recorder, and a special counter with which the absorption curves of low energy electrons may be obtained.

The results confirm a previously published scheme for the main mode of decay, namely that a 400 kev. β -ray from ^{181}Hf leads directly to the 20 μ sec. metastable state in ^{181}Ta . This decays by emission of a 130 kev. γ -ray, followed by a 470 kev. transition to the ground state. Instantaneous coincidences were found to involve the γ -transitions in the main branch, electrons from γ -rays of 87, 133 and 340 kev., and a β -ray of approximately 440 kev.

It is suggested that about 30% of the ^{181}Hf nuclei in the 46 day state decay by a highly forbidden γ -transition (133 kev.) with an 87 kev. γ -ray in cascade to the ground state. This decays through a 440 kev. β -transition to an excited state (340 kev.) of ^{181}Ta . Half life and conversion coefficient values enable spins and relative parities to be assigned to each of the levels occurring in the decay scheme proposed.

On the Development of the Nucleon Component of the Cosmic Radiation in Air, by
H. MESSEL and D. M. RITSON.

ABSTRACT. The development of the nucleon component in air is considered in terms of a model which is an extension of that proposed by Heitler and Jánossy. The numerical results obtained are compared with experiment, and good agreement is found.

SUBJECT INDEX

	PAGE
Aberration, spherical, of magnetic electron lenses, reduction	703
Aberration, spherical correction in electron lenses (Letter)	1037
Aberrations, wave-front of oblique pencils in a symmetrical optical system	709
Abstracts, Cambridge Conference	141
Abstracts—Section A 63, 155, 222, 303, 374, 462, 542, 630, 734, 831, 966,	1051
Alkali metal in glass, preparation (Letter)	455
Alloys, precipitation-hardened, creep in	847
Alloys, uranium and uranium-iron, magnetic properties	520
Amorphous solids, variation with wavelength of refractive index and absolute stress optical coefficients	446
Aqueous sols, electron microscopic studies	980
Arc emission from liquid surface	377
Area of contact and static coefficient of friction	185
Atmosphere and physics	252
Atmospherics, sudden enhancements	122
Aurora and luminous night clouds (Letter)	1039
Breakdown in cold-cathode tubes at low pressure	25
Brightness profile and photometric contrast object silhouetted against twilight sky	364
Calibration of platinum thermocouples over range 0–1,760° C.	492
Cambridge Conference—abstracts	141
Cameras, Schmidt-Cassegrain, monocentric	553
Capacitance measurements on selenium rectifiers: evidence of anomalous dispersion	176
Capacitance of selenium rectifiers (Letter)	1036
Cathode spot, behaviour on undisturbed liquid surface of low work function	377
Cavitation produced by ultrasonics	674
Centimetre wavelengths, properties of ferromagnetic compounds at	65
Charles Chree Address, 5th: Physics and the Atmosphere	252
Cold-cathode tubes, breakdown at low pressure	25
Colour perception in parafoveal vision	83
Colour temperature of light sources	685
Concave space charge flow	1017
Construction of interference filter for the transmission of light of specified wave- lengths	359
Contacts, selenium, electrical forming phenomena at	861
Corrigenda	221
Counters, Geiger, hydrogen-filled	665
Counters, Geiger-Müller, self-quenching, temperature dependence of counter characteristics (Letter)	722
Counters, self-quenching, dead times	15
Coupling of ordinary and extraordinary rays in ionosphere	49
Creep of cadmium crystals, surface effects	339
Creep in precipitation-hardened alloy	847
Creep testing, simple constant stress apparatus	722
Crystal diode and triode action in lead sulphide (Letter)	571
Crystals, cadmium, surface effects in creep	839
Crystals, monoclinic, elastic and piezoelectric coefficients	577
Crystals, <i>see also</i> Single crystals, Zinc crystals.	
Dead times of self-quenching counters	15
Debye effect in electrolytes (Letter)	58
Deformation of mild steel, optical method for study	724
Diffusion, concentration-dependent, influence on rate of evaporation	484

	PAGE
Diodes, retarding field current in	300
Dipole moment measurements, solvent effects	893
Dissociation processes in certain gases of high dielectric strength	180
Elastic coefficients of monoclinic crystals, ethylene diamine tartrate	577
Elastic constants of solid containing spherical holes	2
Electrical forming phenomena at selenium contacts	861
Electrolytes, Debye effect in (Letter)	58
Electron beams, focusing of, in certain magnetic fields	954
Electron lenses, magnetic, reduction of spherical aberration	703
Electron lenses, properties of electrostatic fields in	699
Electron lenses, correction of spherical aberration using correcting foil element (Letter)	1037
Electron microscope, three-stage, electron optics (Letter)	59
Electron microscopy : studies on aqueous sols	980
Electron-optical immersion objective, field in	75
Electron optics of three-stage electron microscope (Letter)	59
Electro-optical shutter, Kerr-cell type, operation and photographic characteristics	561
Electrostatic fields encountered in electron lenses, properties of	699
Elliptical plates, vibrations	451
Energy dissipation by pendulum oscillating in air at low pressures	456
Equivalent circuit of piezoelectric crystal, measurement	323
Evaporation, diffusion-controlled	484
Expansion of gaseous spark channels	649
Experimental studies in thermal convection (33rd Guthrie Lecture)	225
Fading, periodic, of short-wave radio signals	924
Ferromagnetic compounds, properties at centimetre wavelengths	65
Ferromagnetic materials, high frequency permeability of	783
Ferromagnetic materials, high frequency permeability of (Letter)	1040
Ferromagnetics, reversible susceptibility of	1005
Field changes, transient electrostatic fluxmeter for study of	402
Fluxmeter, electrostatic, of short response-time for studying transient field changes	402
Frequency response of PbS transistors (Letter)	540
Friction, static coefficient, and area of contact	185
Frictional relaxation oscillations (Letter)	964
Gallium, liquid, electrical resistance near melting point	662
Gases of high dielectric strength, dissociation processes	180
Geiger counters, <i>see</i> Counters	
Guthrie Lecture, 33rd: Experimental Studies in Thermal Convection.	225
Guthrie Lecture, 34th: Sliding Surface	465
Helium liquefier, Joule-Thomson cascade	504
Holweck Discourse : Weak rays in optical spectra	969
Hydrogen-filled Geiger counters	665
Ice crystals, deposition on cooling surfaces in low temperature plant	157
Igneous rocks, magnetic hysteresis	931
Insulators, good, measurements of resistivity	590
Interference fringes, multiple-beam, formed with curved thin sheets	545
Interferometer, Michelson, at millimetre wavelengths	939
Ionized gases, scattering of 3-cm. radiation by (Letter)	726
Ionized layers, formation, theory	129
Ionized layers, production in non-isothermal atmosphere	427
Ionosphere, coupling of ordinary and extraordinary rays	49
Ionosphere, echo scattering near F_c critical frequency	126
Ionosphere, study of horizontal irregularities	907
Ionospheric cross-modulation : techniques of measurement	616

Subject Index

1055

PAGE

Joule-Thomson cascade liquefier for helium	504
Lattice parameter, precision measurements (Letter)	1034
Lattice parameters of clear crystalline quartz	208
Lead sulphide, crystal diode and triode action (Letter)	371
Light sources, colour temperature of	685
Light transmission, interference filters for.	359
Low temperature plant, deposition of ice crystals on cooling surfaces	157
Magnesium antimonide layers, photoconductivity in	982
Magnetic fields, focusing of electron beams	954
Magnetic hysteresis of igneous rocks	931
Magnetic lens fields, calculation by relaxation methods	386
Magnetic permeability in alnico, time decrease	509
Magnetic properties of uranium and uranium-iron alloys	520
Magneto-resistance of silicon-iron	996
Magnetron, toroidal	278
Magnetrons, multiple, circuit behaviour in neighbourhood of critical anode voltage	41
Meteorology and radio, investigation in South Island of New Zealand	595
Microwave power measurement below 3 cm., discrepancies in (Letter)	215
Microwave power measurement at 3 cm. and 10 cm. (Letter)	623
Monocentric Schmidt-Cassegrain cameras	553
Movements of small ferromagnetic particles in inhomogeneous magnetic fields	12
Multiple-beam interference fringes, localized, formed with curved thin sheets.	545
Night clouds, luminous, and aurora (Letter)	1039
Optical emissivity of titanium and zirconium	573
Optical method for studying deformation of mild steel (Letter)	724
Optical system, symmetrical, wave-front aberrations of oblique pencils in	709
Optical system, symmetrical, wave-front aberrations of oblique pencils (<i>Proc. Phys. Soc. B</i> , 1949, 62, 726), Coing, G.	221
Optical system, weak rays in (Holweck Discourse)	969
Orientated metal films on inorganic single crystals	833
Particles, small ferromagnetic, in inhomogeneous magnetic fields, movements of	12
Pendulum oscillating in air at low pressures, dissipation of energy by	456
Permeability, high frequency, of ferromagnetic materials	783
Permeability, high frequency, of ferromagnetic materials (Letter)	1040
Phase-contrast testing with a slit source	527
Photoconductive cells of cadmium selenide (Letter)	624
Photoconductivity in magnesium antimonide layers	982
Photoconductivity, refractive index and infra-red threshold of sensitivity	157
Photometry of distant object silhouetted against sky	369
Physics and the Atmosphere (5th Charles Chree Address)	253
Piezoelectric coefficients of monoclinic crystals, especially ethylene diamine tartrate	577
Piezoelectric coefficients of quartz, temperature variation	890
Piezoelectric crystals, measurement of equivalent electrical circuit	123
Power measurement, microwave, at 3 cm. and 10 cm. (Letter)	623
Power measurement, microwave, discrepancies in	215
Preparation of alkali metals in glass (Letter)	455
Quantum efficiency in photographic x-ray exposures	90
Quartz, clear crystalline lattice parameters	208
Quartz, piezoelectric coefficients, temperature variation	890
Radio fading, analysis of observation at spaced receivers	106
Radio meteorological investigation in South Island of New Zealand	595
Radio signals, short-wave, periodic fading	924
Rectifiers, selenium, capacitance of (Letter)	1036

	PAGE
Rectifiers, selenium, capacitance measurements : anomalous dispersion	176
Reflection coefficient of water at wavelength 8.7 mm., measurements	46
Refractive index and stress optical coefficients of amorphous solids	446
Rehbinder effect	990
Relaxation methods for calculation of magnetic lens fields	386
Relaxation oscillations, frictional (Letter)	964
Resistance, electrical, of liquid gallium near melting point	662
Resistivity of good insulators, measurements of	590
Resonances, satellite, in ultrasonic interferometry	958
Retarding field current in diodes	300
Scattering of 3-cm. radiation by ionized gases (Letter)	726
Secondary electron emission from molybdenum produced by helium, neon, argon and hydrogen	876
Single crystals of cadmium, thermal etching	198
Single crystals of tin, measurement of the thermal expansion by interferometric method	267
Single crystals, substrates, orientated overgrowths of metal films on	833
Sliding surface (34th Guthrie Lecture)	465
Slipping stream of electrons, wave propagation in	409
Solenoid, 100 kw., water-cooled	1028
Solvent effects in dipole moment measurements	893
Spark channel diameters, measurement of (Letter)	370
Spark channels, gaseous, expansion of	649
Spark channels, gaseous, voltage gradients in	633
Spectra, optical, weak rays in (Holweck Discourse)	969
Stress apparatus for creep testing	346
Surface effects in creep of cadmium crystals	839
Surface effects and plasticity of zinc crystals	331
Surface tensions, in system solid copper-molten lead	350
Susceptibility, reversible, of ferromagnetics	1005
Temperature variation of piezoelectric coefficients of quartz	890
Thermal etching of single crystals of cadmium	198
Thermal expansion of single crystals of tin, measurement of, by interferometric method	267
Thermocouples, E.M.F.-temperature calibration	492
Transistors, PbS, frequency response (Letter)	540
Ultrasonic interferometry, satellite resonances	958
Ultrasonics, cavitation produced by	674
Ultrasonics, summarized proceedings of Symposium	305
Valve theory, congruent space charge flow	1017
Vibrations of free elliptical plates	451
Viscous flow transverse to a circular cylinder	288
Vision, parafoveal, colour perception in	83
Voltage gradients in long gaseous spark channels	633
Water dispersions, fluctuating concentration of x-ray products in	297
Water reflection coefficient measurements at wavelength 8.7 mm.. . . .	46
Wave-front aberrations of oblique pencils in a symmetrical optical system: refraction and transfer formulae (<i>Proc. Phys. Soc. B</i> , 1949, 62, 726), Corrigenda	221
Wave propagation in slipping stream of electrons : small amplitude theory	409
X-ray photographic exposures, quantum efficiency	90
X-ray products in water dispersions, fluctuating concentration	297
Yield points in zinc crystals	339
Zinc crystals, plasticity and surface effects	331
Zinc crystals, yield points in	339

INDEX OF AUTHORS (WITH TITLES)

	PAGE
Adam, G. D., and Standley, K. J. : Ferromagnetic resonance in manganese arsenide (Letter)	820
Andrade, E. N. da C., and Randall, R. F. Y. : The thermal etching of single crystals of cadmium	198
Andrade, E. N. da C., Randall, R. F. Y., and Makin, M. J. : The Reh binder effect	
Ashmead, J. : A Joule-Thomson cascade liquefier for helium	504
Bailey, G. L. J., and Watkins, H. C. : Surface tensions in the system solid copper-molten lead	350
Banbury, P. C., and Henisch, H. K. : On the frequency response of PbS transistors (Letter)	540
Banbury, P. C., <i>see also</i> Gebbie, H. A.	
Barakat, N., <i>see</i> Tolansky, S.	
Barber, C. R. : The E.M.F.-temperature calibration of platinum, 10% rhodium-platinum and platinum, 13% rhodium-platinum thermocouples over the temperature range 0°-1,760° C.	492
Barber, D. R. : Note on the brightness profile and photometric contrast of a test-object having small angular dimensions and silhouetted against the twilight sky	364
Barham, M., <i>see</i> Hopkins, H. H.	
Bates, D. R., and Seaton, M. J. : Theoretical considerations regarding the formation of the ionized layers	129
Bates, J. H., and Mallard, J. R. : The magnetic properties of uranium and uranium-iron alloys	520
Bechmann, R. : Determination of the elastic and piezoelectric coefficients of monoclinic crystals, with particular reference to ethylene diamine tartrate	577
Bell, J. F. W. : Satellite resonances in ultrasonic interferometry	958
Billington, C. : <i>see</i> Greenland, K. M.	
Birks, J. B. : The properties of ferromagnetic compounds at centimetre wavelengths	65
Blaha, F. : On movements of small ferromagnetic particles in inhomogeneous magnetic fields	12
Bradfield, G. : Summarized proceedings of symposium on applications of ultrasonics	905
Bradshaw, F. J. : The topical emissivity of titanium and zirconium	573
Braydon, J. E. H. : A new method of measurement of the variation with wavelength of the refractive index and absolute stress optical coefficients of amorphous solids	446
Briggs, B. H., and Phillips, G. J. : A study of the horizontal irregularities of the ionosphere	507
Briggs, B. H., Phillips, G. J., and Shinn, D. H. : The analysis of observations on spaced receivers of the fading of radio signals	106
Bristow, J. R. : Frictional relaxation oscillations (Letter)	964
Bromley, D., and Herz, R. H. : Quantum efficiency in photographic x-ray exposures	90
Bruckshaw, J. McG., and Rao, B. S. : Magnetic hysteresis of igneous rocks	931
Buneman, O. : A toroidal magnetron	278
Bureau, R. : Les renforcements brusques des ondes très longues	102
Burgess, R. E. : The capacitance of selenium rectifiers (Letter)	1036
Chaghtai, M. S. B. : Note on the electron velocity distribution in low voltage arcs	768
Challice, C. E. : Electron optics of the three-stage electron microscope (Letter)	59
Champion, K. S. W. : The production of pulsed magnetic fields using condenser energy storage	795
Childs, B. G., and Weintraub, S. : The measurement of the thermal expansion of single crystals of tin by an interferometric method	267
Collard, J., Nicoll, G. R., and Lines, A. W. : Discrepancies in the measurement of microwave power at wavelengths below 3 cm. (Letter)	215

	PAGE
Collinge, B. : Dead times of self-quenching counters	15
Collinge, B. : Hydrogen-filled Geiger counters	665
Cooper, R. : Capacitance measurements on selenium rectifiers : evidence of anomalous dispersion	176
Corner, W. D., <i>see</i> Tebble, R. S.	
Cottrell, A. H., <i>see</i> Harper, S.	
Craggs, J. D., <i>see</i> Denno, S. N.	
Craggs, J. D., <i>see</i> Warren, J. W.	
Crank, J. : The influence of concentration-dependent diffusion on rate of evaporation	484
Culshaw, W. : The Michelson interferometer at millimetre wavelengths	939
Daniels, J. M. : A 100-kilowatt water-cooled solenoid	1028
Das, P. M., <i>see</i> Khastgir, S. R.	
Davies, C. N. : Viscous flow transverse to a circular cylinder	288
Davis, M., and Thompson, N. : Creep in a precipitation-hardened alloy	847
Denno, S. N., Prime, H. A., and Craggs, J. D. : The scattering of 3-cm. radiation by ionized gases (Letter)	726
Dobson, G. M. B. : Physics and the atmosphere (5th Charles Chree Address)	252
Dodd, C. : The electrical resistance of liquid gallium in the neighbourhood of its melting point	662
Dodd, C., and Roberts, G. N. : Dielectric loss and dielectric constant measurements in supercooled liquids	814
Driver, K. S., <i>see</i> Webster, F. V.	
Eckersley, T. L. : Coupling of the ordinary and extraordinary rays in the ionosphere	49
von Engel, A., <i>see</i> Francis, G.	
Ewels, J., <i>see</i> Henisch, H. K.	
Finch, G. I. : The sliding surface (34th Guthrie Lecture)	465
Foster, H. G., <i>see</i> Newton, C. C.	
Francis, G., and von Engel, A. : Development of the low pressure electrodeless discharge in a high-frequency electric field (Letter)	823
Froome, K. D. : The behaviour of the cathode spot on an undisturbed liquid surface of low work function	377
Fürth, R., and MacDonald, D. K. C. : On the retarding field current in diodes	300
Gebbie, H. A., Banbury, P. C., and Hogarth, C. A. : Crystal diode and triode action in lead sulphide (Letter)	371
Gianola, U. F. : The correction of the spherical aberration of electron lenses using a correcting foil element (Letter)	1037
Gianola, U. F. : Reduction of the spherical aberration of magnetic electron lenses	703
Gibson, A. F. : The absorption spectra of solid lead sulphide, selenide and telluride	756
Gilbert, Miss M. : Colour perception in parafoveal vision	83
Gledhill, J. A., and Szendrei, M. E. : Theory of the production of an ionized layer in a non-isothermal atmosphere neglecting the earth's curvature, and its application to experimental results	427
Greene, D. : Secondary electron emission from molybdenum produced by helium, neon, argon and hydrogen	876
Greenland, K. M., and Billington, C. : The construction of interference filters for the transmission of light of specified wavelengths	359
Hall, E. O. : An optical method for studying the deformation of mild steel (Letter)	724
Harding, H. G. W. : The colour temperature of light sources	685
Harper, S., and Cottrell, A. H. : Surface effects and the plasticity of zinc crystals	331
Haselden, G. G. : The deposition of ice crystals on cooling surfaces in low temperature plant	157
Hatch, D., <i>see</i> Parker, R. C.	
Hay, H. G., <i>see</i> Marfarlane, G. G.	

Index of Authors (with Titles)

1059

PAGE

Henisch, H. K., and Ewels, J. : A study of electrical forming phenomena at selenium contacts	861
Henisch, H. K., <i>see</i> Banbury, P. C.	
Herz, R. H., <i>see</i> Bromley, D.	
Hesse, Mary B. : The calculation of magnetic lens fields by relaxation methods	386
Heymann, F. G. : Breakdown in cold-cathode tubes at low pressure	25
Higham, J. B., and Meek, J. M. : The expansion of gaseous spark channels	649
Higham, J. B., and Meek, J. M. : Voltage gradients in long gaseous spark channels	633
Higham, J. B., <i>see</i> Saxe, R. F.	
Hogarth, C. A., <i>see</i> Gebbie, H. A.	
Holtham, A. E. J., and Prime, H. A. : The operation and photographic characteristics of a Kerr-cell type of electro-optical shutter	561
Hopkin, L. M. T. : A simple constant stress apparatus for creep testing	346
Hopkins, H. H., and Barham, P. M. : The influence of the condenser on microscopic resolution	737
Hopkins, M. R., and Toye, T. C. : The determination of the viscosity of molten metals	773
Hopwood, W., <i>see</i> Warren, J. W.	
Hunter, A. N. : The Debye effect in electrolytes (Letter)	58
Hyde, F. J., <i>see</i> Newton, C. C.	
Jacob, L. : The field in an electron-optical immersion objective	75
Jacquinet, P. : Quelques recherches sur les raies faibles dans les spectres optiques	969
Keith, H. D. : The lattice parameters of clear crystalline quartz	208
Keith, H. D. : Precision lattice-parameter measurements (Letter)	1034
Kendall, F. F. : Electrical conductivity of gray tin (Letter)	821
Khastgir, S. R., and Das, B. K. : Periodic fading of short-wave radio signals	924
Kiely, D. G. : Measurements of the reflection coefficient of water at a wavelength of 8.7 mm.	46
Krenz, F. H. : On the fluctuating concentration of heavy products in water dispersions	297
Kuhn, H., and Wilson, B. A. : Reflectivity of thin silver films and their use in interferometry	745
Lindsay, P. A. : Certain properties of electrostatic fields encountered in electron lenses	699
Lines, A. W., <i>see</i> Collard, J.	
Linfoot, E. H. : On phase-contrast testing with a slit source	527
Lomar, P. D. : The dielectric strength of aluminium oxide films (Letter)	818
Loosemore, W. R., and Taylor, D. : Temperature dependence of counter characteristics in self-quenching Geiger-Müller counters (Letter)	728
Lynch, A. C. : Measurement of the equivalent electrical circuit of a piezoelectric crystal	323
Lynch, A. C. : The variation with temperature of the piezoelectric coefficients of quartz	890
MacDonald, D. K. C., <i>see</i> Fürth, R.	
MacDonald, D. K. C., and Stanworth, J. E. : Preparation of alkali metals in glass. (Letter)	455
Macfarlane, G. G. : A theory of contact noise in semi-conductors.	807
Macfarlane, G. G., and Hay, H. G. : Wave propagation in a slipping stream of electrons : small amplitude theory.	409
Mackenzie, J. K. : The elastic constants of a solid containing spherical holes.	2
Makin, M. J., <i>see</i> Andrade, E. N. da C.	
Malan, D. J., and Schonland, B. F. J. : An electrostatic fluxmeter of short response-time for use in studies of transient field-changes.	402
Mallard, J. R., <i>see</i> Bates, L. F.	
Meek, J. M., <i>see</i> Higham, J. B.	

	PAGE
Millership, R., and Webster, F. V. : High frequency permeability of ferromagnetic materials.	783
Milnes, B., and Unwin, R. S. : A radio meteorological investigation in the South Island of New Zealand	595
Moss, T. S. : Photoconductivity in magnesium antimonide layers	982
Moss, T. S. : A relationship between the refractive index and the infra-red threshold of sensitivity for photo-conductors.	167
Neppiras, E. A., <i>see</i> Noltingk, B. E.	
Newton, C. C., Hyde, F. J., and Foster, H. G. : Ionospheric cross-modulation : techniques of measurement.	616
Nicoll, G. R., <i>see</i> Collard, J.	
Noltingk, B. E., and Neppiras, E. A. : Cavitation produced by ultrasonics.	674
Parker, R. : A study of the magneto-resistance of silicon-iron	996
Parker, R. C., and Hatch, D. : Frictional relaxation oscillations (Letter)	964
Parker, R. C., and Hatch, D. : The static coefficient of friction and the area of contact.	185
Paton, J. : Aurora and luminous night clouds (Letter)	1039
Phillips, D. J., and Thompson, N. : Surface effects in creep of cadmium crystals.	839
Phillips, G. J., <i>see</i> Briggs, B. H.	
Prime, H. A., <i>see</i> Denno, S. N.	
Prime, H. A., <i>see</i> Holtham, A. E. J.	
Ramsey, N. W. : Some measurements of the resistivity of good insulators.	590
Randall, R. F. Y., <i>see</i> Andrade, E. N. da C.	
Rankine, A. O. : Experimental studies in thermal convection (33rd Guthrie Lecture).	225
Rao, B. S., <i>see</i> Bruckshaw, J. McG.	
Rao, M. R. A. : Electron microscopic studies on aqueous sols.	980
Rivault, R. : Diffusion des échos au voisinage des fréquences critiques de F 2	126
Roberts, G. N., <i>see</i> Dodd, C.	
Robertshaw, R. G., <i>see</i> Willshaw, W. E.	
Ross, I. G., and Sack, R. A. : Solvent effects in dipole moment measurements.	893
Sack, R. A., <i>see</i> Ross, I. G.	
Saxe, R. F., and Higham, J. B. : The measurement of spark channel diameters. (Letter)	370
Schonland, B. F. J., <i>see</i> Malan, D. J.	
Schwarz, E. : Photoconductive cells of cadmium selenide. (Letter).	624
Seaton, M. J., <i>see</i> Bates, D. R.	
Shinn, D. H., <i>see</i> Briggs, B. H.	
Shortt, W. H. : The dissipation of energy by a pendulum oscillating in air at low pressures (Letter)	456
Standley, K. J., <i>see</i> Adam, G. D.	
Stanworth, J. E., <i>see</i> MacDonald, D. K. C.	
Street, R., and Whitaker, P. D. : The measurement of microwave power at wavelengths of 3 cm. and 10 cm. (Letter)	623
Street, R., and Woolley, J. C. : Time decrease of magnetic permeability in alnico	509
Sturrock, P. A. : Note on the focusing of electron beams in certain magnetic fields.	954
Szendrei, M. E., <i>see</i> Gledhill, J. A.	
Taylor, D., <i>see</i> Loosemore, W. R.	
Tebble, R. S., and Corner, W. D. : Investigations on the reversible susceptibility of ferromagnetics.	1005
Thirsk, H. R. : A note on the orientated overgrowths of metal films on single crystal inorganic substrates.	833
Thompson, N., <i>see</i> Davis, M.	
Thompson, N., <i>see</i> Phillips, D. J.	

Tolansky, S., and Barakat, N. : New localized multiple-beam interference fringes formed with curved thin sheets	545
Toye, T. C., <i>see</i> Hopkins, M. R.	
Unwin, R. S., <i>see</i> Milnes, B.	
Wain, H. L., and Cottrell, A. H. : Yield points in zinc crystals.	339
Walker, G. B. : Congruent space charge flow.	1017
Waller, Mary D. : Vibrations of free elliptical plates.	451
Warren, J. W., Hopwood, W., and Craggs, J. D. : On dissociation processes in certain gases of high dielectric strength	180
Watkins, H. C., <i>see</i> Bailey, G. L. J.	
Wayman, P. A. : The monocentric Schmidt-Cassegrain cameras.	553
Webster, F. V., and Driver, K. S. : High frequency permeability of ferromagnetic materials (Letter)	1040
Webster, F. V., <i>also see</i> Millership, R.	
Weinstein, W. : The computation of wave-front aberrations of oblique pencils in a symmetrical optical system.	709
Weintroub, S., <i>see</i> Childs, B. G.	
Whitaker, P. D., <i>see</i> Street, R.	
Willshaw, W. E., and Robertshaw, R. G. : The behaviour of multiple circuit magnetrons in the neighbourhood of the critical anode voltage.	41
Wilson, B. A., <i>see</i> Kufus, H.	
Woodward, G., <i>see</i> Street, R.	

INDEX TO REVIEWS OF BOOKS

	PAGE
Ashworth, J. R. : <i>Ultra-Violet and Daylight Rays</i>	827
Borries, B. von : <i>Die Übermikroskopie</i>	824
Bremmer, H. : <i>Terrestrial Radio Waves : Theory of Propagation</i>	458
Broglie, L. de : <i>Les ondes électromagnétiques centimétriques</i>	828
Brown, T. B. : <i>Foundations of Modern Physics</i>	220
Burdon, R. S. : <i>Surface Tension and the Spreading of Liquids</i>	627
Centre National de la Recherche Scientifique, Paris : <i>Méthodes de calcul dans des problèmes de mécanique</i>	829
Cherry, C. : <i>Pulses and Transients in Communication Circuits</i>	216
Dale, J. B. : <i>Five-figure Tables of Mathematical Functions</i>	152
Dushman, S. : <i>Scientific Foundations of Vacuum Technique</i>	219
Evans, R. M. : <i>An Introduction to Color</i>	220
Ferri, A. : <i>Elements of Aerodynamics of Supersonic Flows</i>	628
Fishenden, M., and Saunders, O. A. : <i>An Introduction to Heat Transfer</i>	824
Francis, V. J. : <i>Fundamentals of Discharge Tube Circuits</i>	151
Fröhlich, H. : <i>Theory of Dielectrics : Dielectric Constant and Dielectric Loss</i>	153
Hartridge, H. : <i>Colours and How We See Them</i>	825
Hartridge, H. : <i>Recent Advances in the Physiology of Vision</i>	828
Henisch, H. K. : <i>Metal Rectifiers</i>	62
Hoyle, F. : <i>Some Recent Researches in Solar Physics</i>	730
Jessop, H. T., and Harris, F. C. : <i>Photoelasticity : Principles and Methods</i>	826
Johnson, B. K. : <i>Optical Design and Lens Computation</i>	151
Keenan, J. H., and Kaye, J. : <i>Gas Tables</i>	826
Magnus, W., and Oberhettinger, F. : <i>Formulas and Theorems for the Special Functions of Mathematical Physics</i>	733
Margenau, H., Watson, W., and Montgomery, C. G. : <i>Physics : Principles and Applications</i>	217
Martin, L. C., and Johnson, B. K. : <i>Practical Microscopy</i>	732
Martin, L. H., and Hill, R. D. : <i>A Manual of Vacuum Practice</i>	626
Merrington, A. C. : <i>Viscometry</i>	217
Moss, H. : <i>Cathode-Ray Tube Traces</i>	9
National Bureau of Standards : <i>Medical X-Ray Protection up to Two Million Volts</i>	41
National Bureau of Standards, Applied Mathematics Laboratory : <i>Tables of Sines and Cosines to Fifteen Decimal Places at Hundredths of a Degree</i>	627
Nightingale, E. : <i>Higher Physics</i>	458
Parodi, M. : <i>Applications physiques de la transformation de Laplace</i>	216
Partridge, J. H. : <i>Glass-to-Metal Seals</i>	829
Peacock, A. G. : <i>Symposium on Electronics</i>	827
Poynting, J. H., and Thomson, J. J. : <i>Sound. A University Text-book of Physics, Volume II</i>	460
Reddick, Harry W. : <i>Differential Equations</i>	733
Rossi, B. B., and Staub, H. H. : <i>Ionization Chambers and Counters : Experimental Techniques</i>	733
Salles, F. T. : <i>Détermination d'un état plan des contraintes à l'aide d'un extensiomètre à résistance électrique à trois directions (rosette) : Abaques pratiques d'emploi</i>	733
Scott Blair, G. W. : <i>A Survey of General and Applied Rheology</i>	372
Sears, F. W. : <i>Principles of Physics : Part III, Optics</i>	372
Sutton, O. G. : <i>Atmospheric Turbulence</i>	731
Trewman, H. F. : <i>Electronics in the Factory</i>	152
U.S. National Applied Mathematics Laboratories : <i>Tables of Scattering Functions for Spherical Particles</i>	218
Weld, L. D. : <i>A Text Book of Heat</i>	453
Wilkes, M. V. : <i>Oscillations of the Earth's Atmosphere</i>	459

